1. Beslissingsondersteuningssystemen voor het Marketing Management* (BOM)

B. WIERENGA

SAMENVATTING

Nadat eerst kort de complexiteit van de marketing-beslissingssituatie is geschetst, wordt in dit artikel een omschrijving gegeven van zogenaamde beslissingsondersteuningssystemen die een belangrijk hulpmiddel kunnen zijn bij een analytische benadering van marketingbeslissingen. Het artikel laat zien hoe in deze systemen marketingmodellen en marketinginformatiesystemen met elkaar worden geïntegreerd, en gaat in op diverse aspecten, zoals het onderscheid in systemen voor strategische versus operationele beslissingen, kwantificering van modellen etc., en geeft een indeling van bestaande modellen. Er worden een aantal concrete ervaringen met deze beslissingsondersteuningssystemen vermeld en het artikel eindigt met een beschouwing over implementatie en de rol die de marktonderzoeker daarin kan spelen.

1. INLEIDING

Marketing management betekent het nemen van beslissingen. Dit kunnen zowel operationele beslissingen zijn: bijvoorbeeld het opstellen van een marketingplan voor een bepaald produkt voor het komende jaar, als strategische beslissingen: het betreden van nieuwe markten, het toevoegen of afstoten van produktielijnen, investeringen in Research & Development enz. Beslissingen kunnen betrekking hebben op één produkt of merk, een groep van produkten, een business unit of de hele onderneming.

Al deze beslissingen hebben als kenmerk dat ze moeten worden genomen in een complexe omgeving met veel onzekere factoren: het koopgedrag van afnemers, het beleid van concurrenten, tussenhandel en de overheid, het verloop van de conjunctuur, de publieke opinie.

De marketingbeslisser zal zijn[1] inzicht in marketingprocessen, zijn kennis van de specifieke markt(en) en zijn ervaring en intuïtie aanwenden om in deze gecompliceerde situatie een marktbeleid uit te stippelen.

Echter in een variant op de titel van een recent boekje over consumentengedrag (Poiesz, 1980): 'De marketingbeslisser is ook maar een mens', die ondanks zijn bekwaamheden beschikt over beperkte vermogens met betrekking tot het analyseren van een situatie, het opnemen en verwerken van informatie, het ontdekken van alternatieven, enz.

In dit artikel wordt nagegaan welke hulp aan de marketingbeslisser kan worden geboden door:

* Artikel naar aanleiding van de voordracht op de NIMA-NVM Themadag over Beslissingsondersteuningssystemen van het Marketing Management op 25 maart 1982 te Wageningen.

BESLISSINGSONDERSTEUNINGSSYSTEMEN[2] VOOR HET MARKETING MANAGEMENT (BOM)

De afnemers en gebruikers van een dergelijk systeem zijn de marketingbeslissers: produktmanagers, marketing managers, marketingdirecteuren, enz. Vanuit een afnemers-gericht denken zullen we in de eerstvolgende paragraaf kort de situatie schetsen waarin marketingbeslissingen genomen worden. Dit geeft een idee hoe een BOM hierin haar bijdrage kan leveren.

Vervolgens wordt aandacht besteed aan de ontwikkelingen die tot deze beslissingsondersteuningssystemen hebben geleid. Daarna worden een aantal specifieke aspecten van een BOM[3] behandeld: het onderscheid in systemen voor strategische en operationele beslissingen, de wijze waarop het marketingproces is weergegeven, het kwantificeren van modellen, een indeling van bestaande modellen, enz. De laatste twee paragrafen bespreken respectievelijk ervaringen met BOM en de problematiek van de implementatie.

2. DE SITUATIE VAN DE MARKETINGBESLISSER

Beslissingen met betrekking tot het marktbeleid worden gekenmerkt door:

- een groot aantal marketinginstrumenten, waaruit een groot aantal verschillende 'mix-en', combinaties van produkt, prijs, promotie en distributie kunnen worden samengesteld;
- interacties tussen de marketinginstrumenten;
- een groot aantal relevante omgevingsvariabelen: afnemers, tussenhandel, concurrentie, overheid, conjunctuur, enz.;
- dynamische effecten waardoor het marktbeleid, gevoerd in een bepaalde periode, invloed heeft op de resultaten en mogelijkheden in toekomstige perioden;
- meerdere doelstellingen: verkopen, omzet, marktaandeel, cash flow, winst, enz.

De marketingbeslisser heeft bescheiden mogelijkheden om hier tegenover te stellen vanwege:

- beperkte beschikbare tijd. Bovendien is de tijd vaak nog versnipperd over een aantal verschillende beslissingsproblemen waarmee men tegelijk bezig is. Volgens Mintzberg (1973) kan het werk van een 'senior manager' worden gekarakteriseerd door 'kortheid, fragmentatie en variatie'. Zelden besteedt een manager langer dan één uur achtereen aan een bepaalde activiteit. Sommige studies maken melding van minstens honderd afzonderlijke activiteiten op één dag (Slevin, 1981);
- beperkte informatie over de voor hem relevante marketingprocessen;
- beperkte vermogens om informatie te verwerken en te analyseren ('computational capacity' Simon, 1979). Het aantal verschillende *alternatieven* (in dit verband kunnen we denken aan het aantal verschillende mix-combinaties of het aantal verschillende marketingstrategieën) die een persoon tegelijk kan overzien, is zeer beperkt. Miller (1956) sprak van zeven, op grond van nader onderzoek kan dit aantal volgens Simon (1979) op vijf

worden gesteld. Daarnaast is het aantal dimensies of *aspecten* waarop een alternatief wordt beoordeeld, beperkt. Bij veel empirisch psychometrisch onderzoek blijkt dit aantal niet meer dan drie of vier te bedragen (Shepard, 1972).

Als resultaat van de gecompliceerdheid van de beslissingssituatie en de bescheiden mogelijkheden die de marketingbeslisser heeft, vertonen de genomen beslissingen veelal de volgende kenmerken:

- Beslissingen zijn in sterke mate op het verleden gebaseerd.
  Eerdere gebeurtenissen en beslissingen uit het verleden vormen vaak het uitgangspunt. Hulbert (1981), in een beschrijvende analyse van marketingbeslissingen, spreekt van 'incremental' gedrag: men doet een klein beetje meer (of minder) dan in het verleden. Bijvoorbeeld iets meer reclame, een kleine prijsstijging e.d..
- Ervaring en intuïtie spelen een belangrijke rol.
- Het beslissingsprobleem wordt vereenvoudigd door slechts enkele alternatieven in beschouwing te nemen en deze te beoordelen op een beperkt aantal aspecten: 'bounded rationality', 'satisficing versus optimizing' (Simon, 1979, p. 3).
- Het hanteren van zg. 'standard-operating rules'. Er zijn beschrijvende modellen gemaakt van de wijze waarop marketingbeslissingen worden genomen: bijvoorbeeld bij prijsstelling (Howard en Morgenroth, 1968), bij produktontwikkeling (Capon en Hulbert, 1975) en bij de beslissing een bepaald produkt wel of niet in het assortiment op te nemen (Montgomery, 1975). In veel gevallen blijkt met een dergelijk model de beslissing goed te kunnen worden voorspeld. Hulbert (1981, p. 44) concludeert dat veel beslissingsregels van een zeer eenvoudig karakter zijn. Een voorbeeld van een dergelijke eenvoudige besluitregel is het veel voorkomende gebruik de reclame-uitgaven vast te stellen als percentage van de omzet (San Augustine en Foley, 1975).

Ondanks de merites van bovengenoemde heuristische beslissingsregels is het allerminst gegarandeerd dat deze zullen leiden tot beslissingen die vanuit de doelstellingen van de organisatie optimaal zijn. Om verder te komen zou de beslisser moeten beschikken over:

- meer informatie;
- analyse van deze informatie, waardoor inzicht wordt verkregen in de werking van het marketing-proces;
- integratie van de verschillende marketinginstrumenten en omgevingsvariabelen in één geheel (een model).

Geprobeerd wordt aan deze behoefte tegemoet te komen met de genoemde beslissingsondersteuningssystemen. Wil een BOM werkelijk effectief zijn, d.w.z. worden gebruikt door de marketingbeslisser, dan moet het kunnen functioneren in de hier beschreven situatie. Dit betekent dat, gezien de beperkte tijd, er zeer snel antwoorden moeten kunnen worden gegeven op gestelde vragen. Dit betekent bovendien dat het systeem moet aansluiten bij het denken van de beslisser, a.h.w. het verlengstuk van zijn denkproces moet zijn: 'Executive Mind Support' (Wagner, 1981).

3. BOM, ANTECEDENTEN EN PERSPECTIEVEN

3.1. *Omschrijving van een BOM*

Naar analogie van de omschrijving die door Keen en Wagner (1979, p. 177) van 'Decision Support Systems' in het algemeen is gegeven, kan een beslissingsondersteuningssysteem voor het *marketing management* worden omschreven als:
een systeem van gegevens en modellen, toegankelijk via een computer, bestemd voor persoonlijk gebruik door de manager of zijn directe staf, dat dient tot directe ondersteuning van zijn marketingbeslissingen.

Belangrijke onderdelen van een BOM zijn (Montgomery, 1974; Little, 1979a):
- gegevens bank;
- modellen;
- statistische procedures;
- optimalisatiemethoden;
- communicatiesysteem met gebruiker.

Een nadere bespreking van deze elementen kan worden gevonden in Meulenberg en Wierenga (1980). Wagner (1981) formuleerde de volgende eisen die aan een BOM moeten worden gesteld:
- snel en gemakkelijk te ontwikkelen en aan te passen;
- snelle response;
- gemakkelijke communicatie met de beslisser;
- de structuur van het systeem moet door de beslisser worden begrepen.

3.2. *Herkomst*

De huidige beslissingsondersteuningssystemen zijn het resultaat van een tweetal ontwikkelingen: die op het gebied van marketingmodellen en van marketinginformatiesystemen.

3.2.1. *Marketingmodellen*

Er bestaat binnen de marketing al een vrij lange traditie om marketingverschijnselen te beschrijven met behulp van wiskundige modellen. Sinds het begin van de jaren zestig zijn er een groot aantal boeken verschenen waarin deze modelmatige aanpak van marketingproblemen centraal staat: Bass e.a. (1961), Frank, Kuehn en Massy (1962), Buzzel (1964), Day (1964), Amstutz (1967), Montgomery en Urban (1969), Stasch (1972), Kotler (1971), Leeflang (1974), Naert en Leeflang (1978). Daarnaast is er een groot aantal artikelen in wetenschappelijke tijdschriften, waarin een modelmatige benadering van marketingproblemen wordt gevolgd. In de eerste periode werd hierbij sterk uitgegaan van de beschikbare optimaliseringsmethoden uit de Operations Research en werd getracht het marketingprobleem zodanig te structureren, dat het paste binnen een standaardmodel. Echter, het bleek dat deze veelal uit de produktieplanning afkomstige methoden voor de marketing weinig realistische oplossingen aandroegen. Het marketingprobleem moest veelal te zeer in een niet-

realistisch wiskundig keurslijf worden geperst. Een sprekend voorbeeld hiervan is de oplossing van het media-planningsprobleem met behulp van lineaire programmering, zoals dat in 1962 werd gedaan door het reclamebureau Batten, Burton, Durstine & Osborn (zie de beschrijving in Buzzel, 1964, Hfd. 5).

Daarna zijn er pogingen gedaan om zeer gedetailleerde totaalmodellen van het hele marketinggebeuren te bouwen. Montgomery en Weinberg (1973) kwalificeren dit als het streven naar 'Bigger and Better'.

Zo ontwikkelde Amstutz (1967) een simulatiemodel voor het hele marketingproces met submodellen voor consumentengedrag, groothandelsgedrag, detaillistengedrag en verkopersgedrag. Een toepassing hiervan is een zeer gedetailleerd model van het voorschrijven van medicijnen door artsen: Claycamp en Amstutz (1968). Stasch (1972) ontwierp modellen voor de subsystemen: vraag, fysieke distributie, persoonlijke verkoop en reclame, en integreerde deze vervolgens tot één totaalsysteem van het marketingproces.

Ondanks het beschikbaar komen van deze steeds meer omvattende modellen, schrijft Little in een artikel in 1970: 'The big problem with management science models is that managers practically never use them.'

Dit min of meer klassiek geworden artikel leidde een beweging in naar meer beslisser-gerichte modellen. Er ontstond een trend naar eenvoudiger, minder ambitieuze modellen die dichter staan bij het beslissingsproces van de manager. Modellen zijn niet in staat de marketingbeslisser te vervangen, maar zullen hem moeten helpen bij het voeren van het marktbeleid. Little (1970, p. B470) komt met het begrip 'decision calculus', omschreven als: 'op modellen gebaseerde procedures voor het verwerken van gegevens en *oordelen*, om de manager te helpen bij zijn beslissingen'. Het opvallende hierbij is dat behalve objectieve gegevens ook het subjectieve oordeel van de manager (dus zijn kijk op de werking van het marketingproces) mee wordt gebruikt in het opstellen en kwantificeren van modellen. Sinds Little's ADBUDG-model (1970) is het decision calculus-principe vele malen toegepast in marketing modellen, o.a. in CALLPLAN (Lodish, 1971), DETAILER (Montgomery e.a., 1971), BRANDAID (Little, 1975) en STRATPORT (Larreché en Srinivasan, 1981).

Ook heeft men in de laatste jaren steeds meer oog gekregen voor het belang van de implementeerbaarheid van een model. Een model dat niet gebruikt wordt door de beslisser, hoe perfect ook in technische zin, is van een zeer beperkte waarde. Aan de problematiek van de implementatie wordt verderop in dit artikel aandacht besteed.

3.2.2. *Marketinginformatiesystemen*

Een afzonderlijke ontwikkeling, die aanvankelijk grotendeels los stond van het gebruik van marketingmodellen, is de opkomst van marketinginformatiesystemen. Hierbij gaat het om het opstellen van 'procedures ten behoeve van een ordelijke stroom van informatie, afkomstig uit bronnen van binnen en buiten de onderneming, ten behoeve van de beslissingen die in de marketing moeten worden genomen' (Brien en Stafford (1968)). Doorgaans speelt de computer hierbij een grote rol. Ook deze ontwikkeling vindt haar oorsprong in de zestiger jaren: Alderson en Shapiro (1963), Hugo (1967), Amstutz (1969). Een goed beeld van het terrein geven de artikelen in Smith, Brien en Stafford (1968).

Het primaire doel van deze informatiesystemen was: verzamelen en opslaan van gegevens en zorgen dat de verschillende beslissers de voor hen relevante informatie krijgen. In de reeks van vragen die men als marketingbeslisser aan het beschikbare materiaal kan stellen:

'Wat is er gebeurd?' (1)
'Waardoor is het gebeurd?' (2)
'Wat gebeurt er, indien...?' (3)
'Wat dient er te gebeuren?' (4)

geeft een aldus gebruikt marketinginformatiesysteem alleen antwoord op de eerste vraag. Echter, hoewel het bijvoorbeeld op zichzelf interessant is te weten dat in een bepaald verkoopdistrict de verkopen in een zekere maand sterk achterblijven, is de marketingman met name geïnteresseerd in de oorzaak van een dergelijk verschijnsel. Is dit een verminderde verkoopinspanning, een prijsnadeel, activiteiten van concurrenten, een dalende totaalmarkt? Voor antwoorden op deze vragen is analyse nodig en moeten samenhangen tussen variabelen worden bestudeerd, waartoe modellen moeten worden opgesteld. Dit geldt ook als men de bovenstaande vragen (3) en (4) wil beantwoorden. Om te kunnen voorspellen wat er gebeurt als bepaalde variabelen veranderen (bijvoorbeeld de eigen marketing-instrumenten) moeten we beschikken over een model dat, hoe globaal eventueel ook, het marketingproces in een markt beschrijft.

Enerzijds zien we dus dat, om de gegevens in een marketinginformatiesysteem te kunnen interpreteren en ze tot hun volle waarde te brengen voor het marketing management, analyse en modellen nodig zijn. Anderzijds heeft men voor marketingmodellen gegevens nodig. Het liefst zo recent mogelijk, opdat het model voor de marketingbeslisser zo relevant mogelijke uitkomsten geeft.
Het is dan ook geen wonder dat marketingmodellen en marketinginformatiesystemen steeds meer naar elkaar toe zijn gegroeid en elkaar hebben gevonden in het BOM. Uiteraard is dit niet bij bominslag gebeurd. In de meer recente beschrijvingen van marketinginformatiesystemen: Montgomery (1974), Versteeg (1980), Meulenberg en Wierenga (1980), vindt men al veel aandacht voor modellen en analysemethoden.

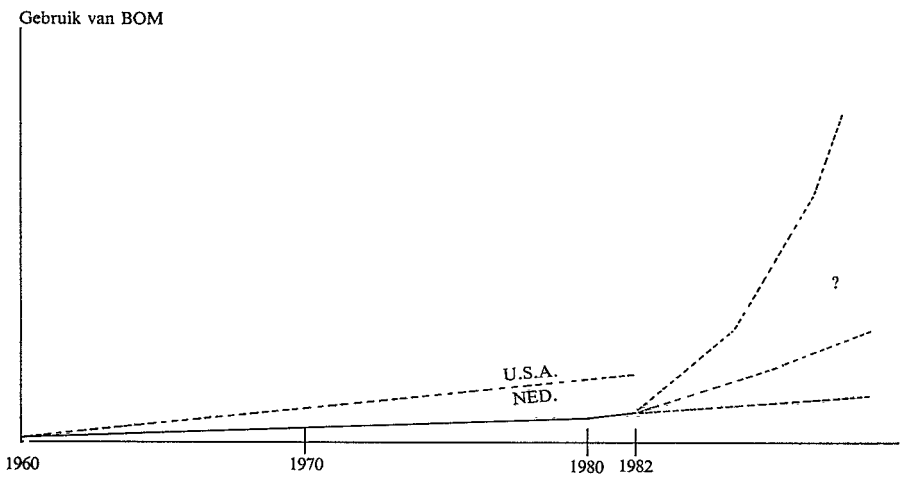
3.3 *Perspectieven*

Het ontstaan van deze geïntegreerde beslissingsondersteuningssystemen wordt begunstigd door de volgende factoren:
(i) Als gevolg van onderzoek op het terrein van de marketing onstaat steeds meer inzicht in de werking van marketinginstrumenten. Dit inzicht is van groot belang bij het opstellen van modellen. Als voorbeeld kan hier worden genoemd het geheel aan inzichten, opgedaan bij empirische studies verricht gedurende de laatste tien jaren, over de relatie tussen reclame-uitgaven en verkopen (Leone en Schultz, 1980).

(ii) Het beschikbaar komen van meer en betere data. Mede als gevolg van automatisering en elektronica kunnen databanken gevoed worden met gegevens uit de eigen onderneming (verkopen, facturering), van detaillistenpanels, con-

sumentenpanels, omnibusonderzoek, scanner-gegevens (UPC), instellingen als CBS, EVD enz. Uiteraard moet er het nodige werk worden verricht om deze gegevens goed met elkaar te laten 'sporen'.

(iii) Ontwikkelingen op het gebied van computers. Op het terrein van de hardware betreft dit vooral zaken als (geheugen)capaciteit, snelheid, gebruikersvriendelijkheid en de fors gedaalde kosten. Belangrijk is ook het beschikbaar komen van minicomputers, die eventueel ook kunnen worden gebruikt als eindstations voor een centrale computer. Hierdoor kan men in een BOM het computergebruik decentraliseren. Verder moeten worden genoemd de toenemende mogelijkheden om over grote afstanden computers aan elkaar te koppelen (telecommunicatie). Bij de software zijn de ontwikkelingen minder spectaculair. De produktiviteitstijging is veel minder dan bij hardware en wordt per jaar op 10% geschat, waardoor naar schatting de software-kosten momenteel 85% van de computerkosten uitmaken (in de jaren vijftig was dit 30%), zie Branscomb (1982). Met name wordt geprobeerd via betere software de computer meer gebruikersvriendelijk te maken, zodat deze toegankelijk is met een minimum aan conventies. Aan de hand van de recente Special over computers en elektronica van het blad Science (1982) kan men zich verder een beeld vormen van de ontwikkelingen op dit terrein.



Figuur 1. Start van de produktlevenscyclus van BOM (speculatief)

Veel van de hier gesignaleerde ontwikkelingen vonden hun oorsprong in de Verenigde Staten. Echter, in Nederland werd aan de rol van de computer in marketing ook al vroegtijdig aandacht besteed. Op 27 november 1970 belegde het NIMA een symposium over 'Marketing en Computer' waar meer dan 400 deelnemers waren. De verwachtingen waren blijkbaar hoog gespannen: 55% van de geënquêteerde aanwezigen verwachtte binnen twee jaren de computer te gebruiken voor modellen in marketing; 84% verwachtte dat de computer ooit in staat zou zijn alle marketingbeslissingen van de mens over te nemen (Marketing en Computer, 1971).

Inmiddels kunnen we, met Little, constateren dat ook in Nederland het gebruik van marketing-modellen nog zeer beperkt is. Er mag echter op grond van bovengenoemde factoren en het meer beslissergericht zijn van BOM, worden aangenomen dat de ontwikkeling een opgaande zal zijn. Welke van de drie in Fig. 1 geschetste paden de produktlevenscyclus van BOM vanaf nu zal inslaan is echter niet met zekerheid te zeggen.

Wel lijkt het waarschijnlijk dat we de bekende S-curve van een zichzelf versterkend proces te zien zullen krijgen. Wanneer een aantal bedrijven zich op het BOM-pad begeven, zullen andere volgen, al was het alleen maar vanwege concurrentie-overwegingen. Immers, een betere analytische basis voor marketingbeslissingen betekent een comparatief voordeel in de markt.

4. SPECIFIEKE ASPECTEN MET BETREKKING TOT BOM

4.1 *BOM voor operationele versus strategische belissingen*

Tot nu toe zijn de ontwikkelde marketingmodellen en marketinginformatiesystemen vooral bestemd geweest voor ondersteuning van operationele marketingbeslissingen: de instelling van de marketing-mix voor een bepaald produkt of een aantal produkten in een bepaalde markt. Er valt echter een toenemende belangstelling te bespeuren voor strategische marketingplanning (Abell & Hammond, 1979, Day, 1977, Wind en Mahajan, 1981) en gezien de betekenis van marketing bij de lange-termijnplanning, dient ook op dit terrein aan systematische dataverzameling en analyse te worden gedaan. Montgomery en Weinberg (1979) beschrijven een zg. Strategisch Intelligence System (SIS), waarmee op een georganiseerde wijze volgens vaste procedures informatie wordt verzameld en opgeslagen met betrekking tot voor strategische planning belangrijke factoren als concurrentie (huidige en toekomstige), technologische ontwikkelingen, de economie, politieke (o.a. overheidsmaatregelen) en sociale factoren. Alleen met behulp van dergelijke informatie is een pro-actief in plaats van een re-actief beleid mogelijk.

Wind (1981) beschrijft de belangrijkste elementen die een marketinggeoriënteerd strategisch planningmodel moet bevatten. Om beslissingen ten aanzien van de toekomstige produkt/markt-portfolio[4] te kunnen nemen (produktontwikkeling, ontwikkeling van nieuwe markten, fusies en overnames), is er een duidelijke integratie nodig van marketing met produktie, financiering, research en development.

Larreché en Srinivasan (1981) komen met een uitgewerkt model voor strategische planning. Ook hier staat het marketingbeleid in de verschillende business units in duidelijk verband met financiële aspecten als investeringen, cash flow, profits e.d. Kenmerkend voor het niveau waarop hier over marketingbeslissingen wordt gedacht, is het gebruik van één functie die voor een bepaalde business unit de relatie tussen het geheel aan marketing-inspanningen ('Marketing Investment') en het marktaandeel beschrijft. Het uitwerken van de totale marketinginspanningen in een meer gedetailleerde marketing-mix is een aangelegenheid van meer operationeel karakter.

4.2. *Mate van gedetailleerdheid waarmee het marketingproces wordt weergegeven*

Veel marketingmodellen hebben een macro-karakter: variabelen als reclameuitgaven en verkopen worden op geaggregeerd niveau met elkaar in verband gebracht, zonder dat wordt gespecificeerd hoe de samenhang op micro-niveau (de consument) werkt. Het model krijgt daardoor een zeker 'black box' karakter, hetgeen zowel uit wetenschappelijk oogpunt als uit het oogpunt van vertrouwenwekkendheid bij de marketing-beslisser nadelig kan zijn. Er valt daarom veel te zeggen voor de ontwikkeling van geïntegreerde modellen waarbij de macro-uitkomsten een gevolg zijn van wat er op micro-niveau gebeurt. Dergelijke modellen geven een veel beter inzicht in het marketingproces. Rao en Sabavala (1981) ontwikkelden zo'n model. Hierin wordt bijvoorbeeld de invloed van de reclame op de verkopen weergegeven door de volgende reeks relaties:
reclame → perceptie van produkten → nut
(preferentie) met betrekking tot produkten → aankoopkansen van produkten → verkopen.

Ook MARKSTRAT (Larreché en Gatignon, 1977), een marketing-game, kent zo'n micro-model voor de werking van marketingvariabelen. Hier veroorzaken produktverandering en reclame gewijzigde posities van de produkten in de meerdimensionale perceptieruimte (andere posities ten opzichte van de constant gezochte ideaalpunten van de consumenten) en langs deze weg hebben de marketingvariabelen invloed op koopkansen en daardoor op marktaandelen. Aldus worden modellen van het consumentengedrag geïntegreerd met marketingbeslissingsmodellen.

4.3. *Kwantificering van de modellen*

Om met een model te kunnen werken moeten de relaties in het model worden gekwantificeerd. Dit betekent bijvoorbeeld dat het model moet aangeven *hoeveel* eenheden van een produkt er *extra* zullen worden verkocht bij een bepaalde toename van de reclame-inspanningen. Een belangrijke bron van gegevens voor deze kwantificering (of parametrisering) vormen de gegevens uit het verleden (zie Leeflang en Reuyl (1982) over het gebruik van dergelijke gegevens voor de parametrisering van een model). Niet altijd zijn dergelijke gegevens echter aanwezig, en soms zijn ze niet (meer) bruikbaar ook al zijn ze beschikbaar. Dit kan worden veroorzaakt door een gebrek aan variatie in het materiaal, maar ook door veranderde omstandigheden waardoor gegevens niet meer relevant zijn: Naert en Weverbergh (1981). Het ontbreken van historische gegevens hoeft echter niet het gebruik van modellen onmogelijk te maken. Genoemd werd al de decision calculus-benadering van Little (1970), waarbij de responsfunctie (de relatie tussen marketing-inspanningen en verkopen) wordt vastgesteld mede op basis van het oordeel van de marketingbeslisser. Op grond van verschillende ervaringen kan worden geconstateerd, dat marketingbeslissingen ook met deze subjectieve elementen bevattende decision calculus-modellen aanmerkelijk kunnen verbeteren (Lodish, 1981). Een belang-

rijk voordeel van deze modellen is dat het noodzakelijk is de marketingbeslisser zeer nauw bij het modelbouwproces te betrekken.

4.4. *Modellen voor beschrijving van het marktmechanisme versus normatieve modellen*

Zoals reeds werd opgemerkt waren de vroege marketingmodellen vooral geïnspireerd door de optimaliseringstechnieken en daardoor sterk normatief gericht. Het ging om het vinden van het optimum: de beste beslissing. De huidige ontwikkeling is meer gericht op het construeren van *modellen van het marktmechanisme* (Parsons, 1981) waarmee de relaties tussen de relevante variabelen in een markt worden beschreven. Met een dergelijk model kan, onder veronderstellingen met betrekking tot de andere variabelen, worden voorspeld hoe het effect van een bepaalde marketing-mix zal zijn op verkopen, marktaandeel, enz., d.w.z. ter beantwoording van de vraag: 'Wat gebeurt er indien?'

Een *normatief model* daarentegen bevat naast een model van het marktmechanisme ook een doelfunctie. Het gaat er dan om een marketing-mix te vinden die deze doelfunctie maximaliseert. Dit is het antwoord op de vraag: 'Wat dient er te gebeuren?' Het formuleren van een passende doelfunctie is niet altijd even eenvoudig: hoe moet bijvoorbeeld korte-termijnwinst worden afgewogen tegen lange-termijn-marktaandeel? Daarnaast leidt een realistische formulering van het marktmechanisme vaak tot modellen die niet rechttoerechtaan met een standaard-optimaliseringstechniek oplosbaar zijn (Zoltners, 1981). Wanneer we een model primair willen laten functioneren als het verlengstuk van het denkproces van de beslisser, dan is de belangrijkste bijdrage de beschrijving van het marktproces. Dit laat de afweging van de resultaten tot het domein van de manager.
Het is met behulp van de huidige computertechniek geen enkel probleem om grote aantallen strategieën door te exerceren (simulatie). Uit deze alternatieven kan vervolgens een keuze worden gemaakt.

4.5. *Indeling van enkele modellen*

Zonder enige pretentie van volledigheid worden in deze paragraaf een aantal in de literatuur verschenen marketingmodellen genoemd. De ruimte voor verdere bijzonderheden over deze modellen ontbreekt. De lezer wordt hiervoor verwezen naar de literatuurreferenties.

Een bruikbare indeling[5] is die naar de fase van de produktlevenscyclus van het produkt. In de sfeer van positionering en design van nieuwe produkten kunnen worden genoemd PERCEPTOR (Urban, 1975) en POSSE (Green, Carroll en Goldberg, 1981). Iets verder in de levenscyclus: het pretestmarktstadium, kunnen de modellen ASSESOR (Silk en Urban, 1978) en SENSOR (Magnin, 1982) worden gebruikt.

Voor de beschrijving van het trial-repeat proces na introductie van een nieuw produkt (eventueel op testmarktschaal) kunnen modellen als SPRINTER (Urban, 1970), DEMON (Learner, 1968) en NEWS (Pringle, Wilson en Brody, 1982) nuttige diensten bewijzen.

Modellen voor het marktbeleid t.b.v. het gevestigde produkt zijn BRANDAID (Little 1975) en MAPLAMOD (Bloom en Stewart, 1977).

Voor de afstervingsfase van een produkt zijn tot nu toe weinig specifieke modellen ontwikkeld. Wel wordt hieraan aandacht besteed in het kader van portfolio-modellen: Day (1977).

Een andere indeling van modellen is naar de instrumenten van de marketingmix waarop ze speciaal zijn gericht. Zo zijn er vrij veel modellen op het terrein van het produktbeleid: naast de hierboven reeds genoemde modellen kan over modelmatige benaderingen op dit gebied nuttige informatie worden gevonden in: Pessemier (1976), Hauser en Urban (1977) en Shocker & Srinivasan (1979). Modellen op het terrein van reclame zijn: ADBUDG (Little, 1970), MEDIAC (Little & Lodish, 1969), ADMOD (Aaker, 1975) en Wierenga (1981). Daarnaast zijn er tal van econometrische 'ad hoc' modellen op het gebied van reclame: voor beschouwingen hierover zie Little (1979b), Leeflang en Reuyl (1982). In de sfeer van verkoopbevordering d.m.v. vertegenwoordigers vinden we de eerder genoemde modellen CALLPLAN (Lodish, 1971) en DETAILER (Montgomery e.a., 1971).

Op het gebied van prijs en distributie zijn de modellen minder talrijk. Voorbeelden zijn: Edelman (1965) voor prijs, en Wierenga en Van Tilburg (1977) voor distributie (detailhandel).

De meeste marketingmodellen hebben een zeker zwaartepunt in één van de mix-elementen. Daarnaast zijn er enkele modellen die in principe de hele marketing-mix beslaan. Voorbeelden hiervan zijn de genoemde modellen BRANDAID en MAPLAMOD.

4.6. *Maatwerk of confectie?*

Uit het voorgaande blijkt dat er in de literatuur een groot aantal modellen zijn verschenen die een min of meer algemene opzet hebben, d.w.z. niet uitsluitend zijn toegesneden op een bepaalde toepassingssituatie. Een aantal van deze modellen wordt door adviesbureaus op de markt gebracht. Niet in alle gevallen kan men een dergelijk model echter zonder meer toepassen. De gebruiker dient zich er terdege van te overtuigen dat de wijze waarop de responsrelaties in het model zijn gespecificeerd, voor zijn situatie passend is. Daarnaast moet worden gewaakt voor het zonder meer overplanten van bepaalde coëfficiënten in het model die in andere markten onder andere omstandigheden zijn geschat (bv. de Amerikaanse markt).

Bij veel modelgebruik is maatwerk geboden, of dienen bestaande modellen op zijn minst te worden aangepast. Hierbij doet men er echter wel verstandig aan niet volledig vanaf nul te beginnen, maar nuttige elementen uit bestaande modellen over te nemen. Op deze wijze kan toch worden geprofiteerd van de grote hoeveelheid ervaring die er inmiddels op het terrein van marketingmodellen is opgedaan.

5. CONCRETE ERVARINGEN MET BOM

Het ontwikkelen en gebruiken van een BOM heeft uiteraard een opbrengsten- en een kostenkant. Wij kunnen hier niet diep ingaan op kosten- en op-

brengstenmeting van marketingmodellen (zie hiervoor Naert & Leeflang, Hfd. 14), maar zullen concreet enkele voorbeelden noemen van toepassingen waarbij de opbrengsten als gevolg van betere belissingen direct aan te geven zijn (soms in concrete geldbedragen).

a. *Prijsstelling: Edelman (1965)*

In dit klassiek geworden artikel gaat het om het afgeven van een prijs voor een bepaalde opdracht, waarvoor men in concurrentie is met andere bieders. De laagste bieder krijgt de opdracht. De verwachtingen m.b.t. de hoogte van de concurrerende prijzen zijn essentieel. Edelman laat zien dat het gebruik van een model waarbij de kwantificering van dit model volledig gebaseerd is op het *oordeel* van de manager, tot systematisch betere prijzen leidt (vaker onder het laagste concurrerende bod en met een kleiner verschil tussen dit bod en het eigen bod) dan wanneer de managers zonder model een prijs afgeven.

b. *CALLPLAN: Lodish (1971, 1981)*

In dit model voor de planning van vertegenwoordigersbezoeken moeten de vertegenwoordigers zelf de responsfuncties aangeven: het effect van het aantal bezoeken op de verkopen aan een klant. CALLPLAN is inmiddels 25 keren toegepast (Lodish, 1981).

Dit model is één van de weinige waarbij door middel van een veldexperiment de invloed van modelgebruik op het resultaat is vastgesteld. Fudge en Lodish (1977) beschrijven een experiment met tien paren vertegenwoordigers voor een luchtvaartmaatschappij, waarbij van ieder paar één (aselect bepaalde) vertegenwoordiger met CALLPLAN werkte en de ander niet. De stijging van de verkopen in de zes maanden durende onderzoekperiode was bij de CALLPLAN gebruikers 8% groter dan bij de vertegenwoordigers die niet met dit model hadden gewerkt.

c. *Andere Decision Calculus-modellen (Lodish, 1981)*

Lodish beschrijft de opzet en implementatie van een BOM voor een detailhandelsorganisatie. De door het BOM veroorzaakte meer analytische benadering van de marketingbeslissingen maakte duidelijk waar belangrijke lacunes zaten met betrekking tot kennis van de werking van de markt. Dit leidde tot het uitvoeren van experimenten, o.a. met betrekking tot prijsstelling en mediakeuze bij reclame. Vermeld wordt dat alleen al de winstverbetering als gevolg van beleidsaanpassingen op grond van deze experimentresultaten vele malen groter is dan de kosten van het bouwen van het systeem.

Lodish beschrijft verder de toepassing van BRANDAID in een onderneming van consumentenprodukten. Nadat BRANDAID was geïmplementeerd, met veel subjectieve input van de betrokken managers, werd het model gebruikt voor het doorrekenen van alternatieve strategieën. De meest perspectief biedende strategie, een prijsverhoging, werd doorgevoerd en leidde tot een winstverbeteringscijfer in de orde van grootte van $5 miljoen.

d. *MAPLAMOD (Bloom en Stewart, 1977)*

Het model van Bloom en Stewart, ontwikkeld voor Beecham Products (frisdranken, toiletartikelen, pharmaceutica) heeft een zeer eenvoudige structuur,

waarbij de marketingvariabelen prijs, reclame en below-the-line activiteiten, in het model zijn opgenomen. Het model heeft het voordeel dat het direct aansluit bij de wijze waarop jaaroverzichten en jaarplannen voor individuele merken worden opgesteld. Hierdoor wordt het model vooral gebruikt bij de bepaling van jaarbudgetten voor reclame en below-the-line. Ook hier is een deel van de input subjectief. Het model wordt interactief gebruikt met behulp van een terminal. Behalve voor de jaarlijkse budgettering wordt het model ook gebruikt voor speciale analyses. Zo werd bijvoorbeeld met behulp van het model ontdekt dat een produkt dat aan het einde van de levenscyclus scheen, via een forse prijsdaling weer winstgevend kon worden gemaakt.

e. *Een reclamemodel (Wierenga, 1981)*

Dit model beschrijft de invloed van reclame-activiteiten in verschillende media (radio, T.V., dagbladen) op de bezoekersaantallen van een recreatiepark. Hierbij gaat het zowel om de verdeling van het reclamebudget over de media en in de tijd (de bepaling van de tijdstippen van de reclame-uitingen) als om het vraagstuk van de hoogte van het reclamebudget.

Het model werd gekwantificeerd op basis van historische gegevens. Het bleek dat door een betere allocatie van het huidige reclamebudget het aantal bezoekers per jaar met 110.000 kan worden verhoogd. T.V.- en radioreclame zou moeten worden opgevoerd ten koste van dagbladen. Verder bleek dat de marginale opbrengst van de reclamegelden op basis van daarmee gegenereerde bezoekersaantallen ongeveer twee gulden per gulden reclame bedroeg. Er was dus zeker geen sprake van 'overspending', maar eerder aanleiding tot enige opvoering van het reclamebudget.

f. *Econometrische modellen in het algemeen*

Bass (1980) beschrijft enkele gevallen van het gebruik van econometrische modellen in marketing en gaat in het bijzonder in op het winstpotentieel van de toepassing van deze modellen bij reclame.

g. *Laboratoriumexperimenten*

Behalve op praktijkschaal zijn sommige modellen ook in het laboratorium beproefd. McIntyre (1982) rapporteert over een experiment met MBA-studenten die met behulp van een CALLPLAN-model een promotioneel budget over territoria moesten verdelen. Een andere even grote groep moest dezelfde verdeling uitvoeren zonder model. Het bleek dat de modelgebruikers een hogere winst realiseerden en minder grote fouten maakten in de toewijzing van de budgetten. Een ander opvallend verschijnsel was dat modelgebruikers grotere stappen durfden te nemen (grotere veranderingen in de beslissingen) van de ene op de andere periode. Waarschijnlijk hangt dit samen met het grotere vertrouwen in hun beslissingen die bij modelgebruikers werd geconstateerd. De resultaten van Chakravarti e.a. (1981), die ook onder laboratoriumomstandigheden beslissers met een model lieten werken (ADBUDG), laten geen positief effect van het model zien. Waarschijnlijk moeten deze resultaten echter worden toegeschreven aan specifieke omstandigheden met betrekking tot het model (Little en Lodish, 1981).

6. IMPLEMENTATIE VAN BOM

6.1. *Voorwaarden voor implementatie:*

'A good many marketing executives in the deepest recesses of their psyches are artists, not analysts. .....Their temperament is antipathetic to system, order, knowledge..... They revel in chaos, abhor facts and fear research', Adler (1967, p. 86).

Een BOM kan alleen succesvol zijn als het gebruikt wordt en werkelijk van invloed is op de genomen marketingbeslissingen. Daarom wordt in toenemende mate bij 'Decision Support Systems' aandacht besteed aan implementatie: Schultz en Henry (1981), Naert en Leeflang (1978, Hfd. 13). De communicatie tussen gebruiker (de marketingbeslisser) en de modelbouwer krijgt hierbij veel aandacht. Eerder in dit artikel werd er al op gewezen dat bij de opzet van een BOM rekening moet worden gehouden met de situatie waarin de marketingbeslisser opereert. Hoewel het in de aanhef van deze paragraaf gegeven citaat de analytische instelling van de marketingbeslisser wellicht al te negatief afschildert, moet toch rekening worden gehouden met een wat andere mentaliteit dan die van de modelbouwer: 'Marketing managers tend to act before they think, if they ever think; management scientists think before they act, if they ever act' (aangepast naar Little, 1970).

Voor het aangeven van de mate van wederzijds begrip tussen modelbouwer en modelgebruiker wordt vaak de indeling van Churchman en Schainblatt (1965) gehanteerd. De minst ideale situatie is die waarbij noch de modelbouwer de modelgebruiker begrijpt, noch omgekeerd. Tussenposities zijn dat er in de ene richting wel begrip is, maar in de andere richting niet. Het beste is uiteraard de situatie waarin de modelbouwer een goed begrip heeft voor de problematiek waar de marketingbeslisser voor staat, terwijl anderzijds de marketingbeslisser (althans in grote lijnen) weet hoe de gebruikte modellen in elkaar zitten.

In de 'traditionele situatie' werken de beslisser en de modelbouwer geheel gescheiden. Op een bepaald moment is het model klaar en wordt de beslisser geacht het model te gebruiken. Het is allerminst zeker dat dit dan ook gebeurt. Een 'evaluationair' modelbouwproces, waarbij met een erg simpel model wordt begonnen, dat vervolgens in samenspraak met de beslisser verder wordt uitgebouwd, biedt veel meer kans op succes (Urban en Karash, 1971; Alavi en Henderson, 1981). Participatie van de beslissers in het modelbouwproces is erg belangrijk.

Als andere factoren die belangrijk zijn voor de geslaagde implementatie van een BOM kunnen worden genoemd (Schultz en Henry, 1981):

- 'organisational validity': mate waarin het BOM past in de bestaande organisatie;
- congruentie van doelstellingen van modelbouwer en modelgebruiker;
- persoonlijk belang van de beslisser bij modelgebruik;
- beslissingsstijl of cognitieve stijl van de beslisser (meer analytisch versus meer heuristisch ingestelde individuen);
- mate van dringendheid;

– neiging tot verandering;
– persoonlijke relatie tussen marketingbeslisser en modelbouwer;
– steun van de topleiding.

6.2. *De rol van de marktonderzoeker m.b.t. BOM*

Tot nu toe heeft de marketingbeslisser in dit artikel centraal gestaan. Immers, ter ondersteuning van zijn beslissingen is het BOM bedoeld. Welke is echter de rol van de marktonderzoeker in dezen? Minimaal moet hij er rekening mee houden dat de door hem verzamelde gegevens ingevoerd zullen moeten worden in het BOM en daarom qua karakteristieken, zoals gebruikte indelingen in produkten, periodenindelingen e.d. met de andere gegevens in het BOM moeten sporen.

Een veel aantrekkelijker situatie lijkt echter die waarbij de marktonderzoeker een centrale rol krijgt bij de opzet en het functioneren van het BOM.

Dit betekent dat men verder gaat dan de traditionele rol van het aandragen van door de belissers gevraagde informatie over markten. Heel nadrukkelijk gaat de marktonderzoeker zich dan bezighouden met de rol die de verschillende soorten informatie spelen in de te nemen marketingbeslissingen. Vanuit deze oriëntatie zal hij initiatieven nemen tot uitbouw/aanpassing van het BOM, zodat dit in de toekomst zijn beslissingsondersteunende rol nog beter zal kunnen vervullen. De marktonderzoeker kan aldus als 'informatie- en beslissingsanalist' de directe verbinding vormen tussen de marketingbeslisser en het BOM.

Voor de externe marktonderzoeker liggen er mogelijkheden in het opbouwen van expertise op het terrein van beslissingsondersteuningssystemen. Met name de kleine en middelgrote ondernemingen zullen in het algemeen niet zelf de know-how in huis hebben voor het opzetten van een BOM. Deze deskundigheid moet dus worden ingekocht. Enerzijds kan deze afkomstig zijn van computerspecialisten die zich oriënteren op de marketing, anderzijds kunnen marktonderzoekbureaus met deskundigheid inzake computers, modellen, etc. deze adviezen geven.

Zowel voor de interne als de externe marktonderzoekers zal de opbouw van deskundigheid ten behoeve van BOM de nodige studie vragen. Deze zal liggen in de sfeer van informatica, systeemanalyse, programmeren, modellenbouw en statistiek. We zagen eerder dat sofware-ontwikkeling bij het opzetten van een BOM een zeer belangrijke activiteit is. Aan deze scholing zal in de komende tijd de nodige aandacht moeten worden besteed. Op deze wijze kunnen, in onderling samenspel, marketingbeslissers en marktonderzoekers komen tot een zo goed mogelijke opzet en gebruik van BOM, steeds gericht op het uiteindelijke doel: betere (= meer profijtelijke) beslissingen in het marketing management.

NOTEN

1. In deze tekst wordt de marketingbeslisser steeds in de mannelijke vorm aangeduid. Uiteraard kunnen de desbetreffende functies evengoed door vrouwen als door mannen worden uitgeoefend.
2. Engelse term. 'Decision Support Systems'.
3. De afkorting BOM kan staan zowel voor het meervoud: Beslissings Ondersteuningssyste*men* als voor het enkelvoud: Beslissings Ondersteuningssys*teem*.
4. Als een onderneming met verschillende produkten actief is op verschillende markten, dan wordt het geheel aan produkt/markt-combinaties van die onderneming een 'portfolio' genoemd. Voor literatuur over portfolio-analyse zij verwezen naar: Abell en Hammond (1979), Day (1977) en Wind en Mahajan (1981).
5. Eerder werd een indeling gemaakt in strategische modellen versus modellen voor operationele beslissingen. Van de eerst gevonden categorie werden er een aantal genoemd. De hier genoemde modellen zijn vrijwel alle gericht op operationele beslissingen.

BIBLIOGRAFIE

Aaker, D.A., 1975, 'ADMOD: An Advertising Decision Model', *J. of Mark. Res.*, Vol. XII (Febr.) 37-45.

Abell, D.F. en J.S. Hammond (1979), *Strategic Market Planning,* Prentice Hall, Englewood Cliffs N.J., 527 p.

Adler, L., 1967, 'Systems Approach to Marketing', *Harv. Bus: Rev.*, 45, May-June, 105-118.

Alavi, M. en J.C. Henderson, 1981, 'Evolutionary Strategy for Implementing a Decision Support System', *Manag.Sc.*, 27, (Nov.), 1309-23.

Alderson, W. en S.J. Shapiro (eds.), 1963, *Marketing and the Computer,* Prentice Hal, Englewood Cliffs N.J., 443 p.

Amstutz, A.E., 1967, *Computer Simulation of Competitive Market Response,* MIT-Press, Cambridge Mass., 457 p.

Amstutz, A.E., 1969, 'Market-Oriented Management Systems: The Current Status', *J. of Mark. Res.*, Vol. VI, (Nov.), 481-96.

Bass, F.M. e.a., (eds.), 1961, *Mathematical Models and Methods in Marketing,* Irwin, Homewood, Ill. 545 p.

Bass, F.M., 1980, 'Some Case Histories of Econometric Modeling in Marketing: What Really Happened', *Interfaces,* 10 (Febr.) 86-90.

Bloom, D. en M.J. Stewart, 1977, 'An Integrated Marketing Planning System', *ESOMAR Seminar* en *Marketing Management Information Systems,* Febr. 2-4, Brussels, p. 168-185.

Branscomb, L.M., 1982, 'Electronics and Computers: An Overview', *Science,* 215, no 4534, 12 February, 755-760.

Brien, R.H. en J.E. Stafford, 1968, 'Marketing Information Systems: A New Dimension for Marketing Research', *J. of Mark.*, 32, (July).

Buzzel, R.D., 1964, *Mathematical Models and Marketing Management,* Division of Research, Grad. School of Buss. Adm., Harvard Univ., Boston, 281 p.

Capon, N. en J.M. Hulbert, 1976, 'Decision Systems Analysis in Industrial Marketing', *Ind. Mark. Manag.*, 4, 143-160.

Chakravarti, D., A. Mitchell en R. Staelin, 1981, 'Judgment bases marketing decision models: problems and possible solutions', *J. of Mark.*, 45, (Fall), 13-23.

Churchman, C. W. en A. H. Schainblatt, 1965, 'The Researcher and the Manager: A Dialectic of Implementation', *Manag.Sc.*, 11 (Febr.), B69-87.

Claycamp, H.J. en A.E. Amstutz, 1968, 'Simulation Techniques in the Analyses of Marketing Strategy', in F.M. Bass e.a. (eds.): *Applications of the Sciences in Marketing Management,* Wiley, New York.

Day, G.S., 1977, 'Diagnosing the Product Portfolio', J. of Mark., 41 (April) 29-38.

Day, R.L., (ed.), 1964, *Marketing Models, Quantitative and Behavioral,* Internat. Textbook Comp., Scanton Penn, 671 p.

Edelman, F. 1965, 'Art and Science of Competitive Bidding', *Harv.Bus.Rev.*, (July-Aug.), 53-66.

Frank, R.E., A.A. Kuehn en W.F. Massy (eds.) 1962, *Quantitative Techniques in Marketing Analyses,* Irwin, Homewood Ill, 556 p.
Fudge, W.K. en L.M. Lodish, 1977, Evaluation of the Effectiveness of a Model Based Salesman's Planning System by Field Experimentation', *Interfaces,* 8, 97-106.
Green, P.E., J.D. Carroll en S.M. Goldberg, 1981, 'A General Approach to Product Design Optimization Via Conjoint Analysis', *J. of Mark.,* 45 (Summer), 17-37.
Hauser, J.R. en G.L. Urban, 1977 'A Normative Methodology for Modeling Consumer Response in Innovation', *Oper.Res.,* 26 (May-June), 406-421.
Howard, J.A. en W.M. Morgenroth, 1968, 'Information Processing Model of Executive Decisions', *Manag.Sc.,* 14 (March) 416-428.
Hugo, I. St.J., 1967, *Marketing and the Computer,* Pergamon Press, Oxford.
Halbert, J.M., 1981, 'Descriptive Models of Marketing Decisions', in R.L. Schultz en A.A. Zoltners (eds.) *Marketing Decision Models,* North Holland, New York, 19-53.
Keen, P.G. en G.R. Wagner, 1979, 'DSS: An Executive Mind-Support System', *Datamation,* (Nov.), 117-122.
Kotler, P., 1971, *Marketing Decision Making: A Model Building Approach,* Holt, Rinehart & Winston, New York, 720 p.
Larreché, J.C. en H. Gatignon, 1977, *MARKSTRAT: A Marketing Strategy Game,* The Scientific Press, Palo Alto Ca.
Larreché, J.C. en V. Srinivasan, 1981, STRATPORT: A Decision Support System for Strategic Planning, *J. of Mark.,* 45 (Fall), 39-52.
Learner, D.B., 1968, 'Profit Maximization through New-Product Marketing Planning and Control', in F.M. Bass e.a. (eds.): *Applications of the Sciences in Marketing Management,* Wiley, New York, 151-167.
Leeflang, P.S.H., 1974, *Mathematical Models in Marketing* Stenfert Kroese, Leiden, 213 p.
Leeflang, P.S.H. en J.C. Reuyl, 1982, 'De bepaling van de effectiviteit van marketinginstrumenten: mogelijkheden, problemen en illustraties', *Tijdschrift voor Marketing* (Juni).
Leone, R.P. en R.L. Schultz, 1980, A study of marketing Generalization, *J. of Mark.,* (Winter), 10-18.
Little, J.D.C., 1970, 'Models and Managers: the concept of a decision calculus', *Manag. Sc.,* 16 (April), B-466-85.
Little, J.D.C., 1975, BRANDAID: A Marketing Mix Model, Parts I and II, *Oper. Res.,* (July-Aug.), 628-673.
Little, J.D.C., 1979a, 'Decision Support Systems for Marketing Managers', *J. of Mark.,* 43, (Summer), 9-26.
Little, J.D.C., 1979b, 'Aggregate Advertising Models: The State of the Art, *Oper. Res.,* 27, (July-Aug.), 629-667.
Little, J.D.C. en L.M. Lodish, 1969, 'A Medium Planning Calculus', *Oper. Res.,* 17, 1-35.
Lodish, L.M., 1971, CALLPLAN: An Interactive Salesman: Call Planning system, *Manag. Sc.,* 18, no. 4, Part II, 25-40.
Lodish, L.M., 1981, 'Experience with Decision-Calculus Models and Decision Support Systems', in R.L. Schultz en A.A. Zoltners (eds.): *Marketing Decision Models,* North Holland, New York, 165-182.
Magnin, L.C. 1982, 'SENSOR, een marketing mix test voor het nieuwe product', *Tijdschrift voor Marketing* (Juni).
Marketing en Computer, 1971, Verslag Symposium van 27 november 1970, NIVE/Stenfert Kroese, 76 p.
Mc. Intyre, S., 1982, 'Experimental Study of the Impact of Judgment-Based Marketing Models', *Manag. Sc.,* 28, (January), 17-33.
Meulenberg, M.T.G. en B. Wierenga, 1980, 'Gebruiker-georiënteerde informatiesystemen voor marketing beslissingen', *Tijdschrift voor Marketing* (Juni), 3-10.
Miller, G.A., 1956, The magical number seven plus or minus two: Some limits on our capacity for processing information, *Psychological Rev.,* 63, 81-97.
Mintzberg, H., 1973, *The Nature of Managerial Work,* Harper & Row, New York.
Montgomery, D.B. 1974, 'Marketing Decision Information systems: Some Design Considerations', in: R. Ferber (ed.): *Handbook of Marketing Research,* 1-44 tot 1-62.
Montgomery, D.B. 1975, New Product Distribution – An Analysis of Supermarket Buyer Decisions, *J. of Mark. Res.,* 12, (August), 255-64.

Montgomery, D.B., A.J. Silk en C.C. Zaragoza, 1971, 'A Multiproduct Sales Effort Allocation Model', *Manag. Sc.,* 18, (December), 3-24.
Montgomery, D.B. en G.L. Urban, 1969, *Management Science in Marketing,* Prentice Hall, Englewood Cliffs, N.J.
Montgomery, D.B. en C.B. Weinberg, 1973, 'Modeling marketing phenomena: a managerial perspective', *J. of Contemporary Business,* (Autumn), 17-43.
Montgomery, D.B. en C.B. Weinberg, 1979, 'Toward Strategic Intelligence Systems', *Journal of Marketing,* 43, (Fall), 41-52.
Naert, P.A. en P.S.H. Leeflang, 1978, *Building Implementable Marketing Models,* Martinus Nijhof, Leiden, 406 pag.
Naert, P.A. en M. Weverbergh, 1981, 'Subjective Versus Empirical Decision Models', in R.L. Schultz en A.A. Zoltners (eds.): *Marketing Decision Models,* North Holland, New York, 99-123.
Parsons, L.J. 1981, 'Models of Market Mechanisms', in R.L. Schultz en A.A. Zoltners (eds.): *Marketing Decision Models,* North Holland, New York, 77-98.
Pessemier, E.A., 1976, *Product Management: Strategy and Organisation,* Wiley, New York.
Poiesz, T., 1980, *Een konsument is ook maar een mens,* Spectrum, Utrecht, 223 p.
Pringle, L., R.D. Wilson en E. Brody, 1982, NEWS: A Decision Oriented Model for New Product Analysis and Forecasting, *Marketing Science,* Vol. 1, no. 1.
Rao, V.R. en D.J. Sabavala, 1981, Some Issues in the Construction of Models for Marketing Decisions, *Research in Marketing,* 4, 251-272.
San Augustine, A.J. en W.F. Foley, 1975, 'How Large Advertisers Set Budget,', *J. of Advertising Research,* 15, (October), 11-16.
Science, 1982, Special Issue on Computer and Electronics, 215, no. 4534, 12 February.
Schultz R.L. en M.D. Henry, 1981, 'Implementing Decision Models,' in R.L. Schultz en A. A. Zoltners (eds): *Marketing Decision Models,* North Holland, New York, 275-296.
Shepard, R.N., A.K. Romney en S.B. Nerlove (eds.) 1972, *Multidimensional Scaling, Theory and Application in the Behavioural Sciences,* Vol. I, Theory, Seminar Press, New York.
Shocker, A.D. en V. Srinivasan, 1979, Multiattribute approaches for product concept evaluation and generation: a critical review, *J. of Mark. Res.,* 16, 159-180.
Silk A.J. en G.L. Urban, 1978, Pre-Test Market Evaluation of News Packaged Goods: A Model and Measurement Methodology, *J. of Mark. Res.,* XV, (May), 171-191.
Simon, H.A., 1979, *Models of thought,* Yale University Press, New Haven, 524 p.
Slevin, D.P., 1981, 'Marketing Models and Organizational Design', in: R.L. Schultz en A.A. Zoltners (eds.): *Marketing Decision Models,* Holland, New York, 1-18.
Smith, S.V., R.H. Brien en J.E. Stafford (eds.), 1968, *Readings in Marketing Information Systems,* Houghton Mifflin, Boston, 399 p.
Stasch, S.F., 1972, *Systems Analysis for Marketing Planning and Control,* Scott, Foresman & Comp, Glenview, Ill, 532 p.
Urban, G.L., 1970, 'SPRINTER Mod. III: A Model for the Analysis of New Frequently Purchasid Consumer Products', *Oper. Res.,* 18, 805-854.
Urban, G.L., 1965, PERCEPTOR: A Model for Product Positioning, *Manag. Sc.,* 21, 858-871.
Urban, G.L. en Karash, 1971, 'Evolutionary Model Building', *J. of Mark. Res.,* 8, 62-66.
Versteeg, J.W., 1980, 'Het Marketing Informatie Systeem', *Tijdschrift voor Marketing,* 14, (Mei).
Wagner G.R., 1981, 'Decision Support Systems: The Real Substance', *Interfaces,* 11, (April), 77-86.
Wierenga, B., en A. van Tilburg, 1977, 'Determining the optimal policy for a home delivery milk retailing: A simulation approach', *European Rev. of Agric. Econ.,* 4, 245, 270.
Wierenga, B., 1981, 'Modelling impact of advertising and optimizing advertising policy: an application in recreation', *European Journal of Operational Research,* 8, 235-248.
Wind, Y., 1981, 'Market Oriented Strategic Planning Models', in R.L. Schultz en A.A. Zoltners (eds.): *Marketing Decision Models,* North Holland, New York, 207-250.
Wind, Y. en V. Mahajan, 1981, 'Designing Product and Business Portfolios', *Harv. Bus. Rev.,* 59, (Jan.-Febr.), 155-165.
Zoltners A.A., 1981, 'Normative Marketing Models', in R.L. Schultz en A.A. Zoltners (eds.); *Marketing Decision Models,* North Holland, New York, 55-76.

2. Enquêteren met de computer: een nieuwe kans voor 'moeilijke' vraagtechnieken

A.E. BRONNER EN R. DE HOOG

SAMENVATTING

Vele jaren werden computers in het sociaal-wetenschappelijk onderzoek alleen voor de analyse van gegevens gebruikt. De laatste jaren breiden de toepassingsmogelijkheden zich ook uit tot de dataverzamelingsfase. Voor- en nadelen hiervan worden geschetst. Het aantrekkelijke van de computer is dat deze dataverzamelingstechnieken mogelijk maakt die in een gewone enquête vanwege de complexiteit niet of nauwelijks hanteerbaar zijn. Dit betekent enerzijds nieuwe kansen voor oude technieken, zoals conjunct meten en de repertory grid, en anderzijds de mogelijkheid technieken te ontwikkelen die voordien nog niet bestonden. Als voorbeeld van een nieuw ontwikkelde techniek wordt het programma GODDESS gepresenteerd, dat wellicht in de toekomst uitgebouwd kan worden tot een programma voor produktontwikkeling. Kort geformuleerd is het kernvoordeel van al deze technieken dat dynamisch geïnterviewd kan worden, dat wil zeggen dat de volgende vraag die gesteld wordt afhankelijk is van het antwoord op de vorige vraag. Deze wijze van interviewen kan vergeleken worden met schaken met een computer. Aan het slot van het artikel worden problemen geschetst, zoals het ondervangen van onverwachte gebruikersacties, en begrijpelijkheid van formuleringen.

1. INLEIDING

In het sociaal-wetenschappelijk onderzoek heeft de computer zich vanaf het midden van de zestiger jaren een duidelijke plaats verworven in die fase van het onderzoek, waarin de gegevens verwerkt worden. Aanvankelijk lag de nadruk op het gebruik van ponskaarten en vaak door de onderzoeker zelf geschreven ad hoc programma's, later verschoof het accent naar complete programmapakketten, voorzien van een eigen stuurtaal (SPSS, BMDP e.d.). Toch bleef in deze verschuiving één ding hetzelfde: de computer werd vooral gebruikt voor rekenwerk, voor het *analyseren* van gegevens. Pas de laatste jaren zijn nieuwe toepassingen ontdekt. Deze ontwikkeling hing samen met de volgende factoren:

- verbeterde mogelijkheden voor interactief gebruik van de computer, dat wil zeggen dat het niet meer strikt noodzakelijk is gegevens eerst via de ponskaart in de computer te brengen. Dit leidde tot zogenaamde *on-line* dataverwerking;
- de mogelijkheid om via telefoonlijnen ook op afgelegen plaatsen met de computer in contact te komen, hetgeen in feite leidde tot *mobiliteit* van de computer
- de *afnemende kosten* van computer hardware, vooral belichaamd in de opkomst van de mini- en micro-computers. Ook deze micro-computers verhogen de mobiliteit.

Gezamenlijk hadden deze factoren als gevolg dat geëxperimenteerd kon worden met het enquêteren via de computer. Een parallelle ontwikkeling viel ook te constateren in de medische wetenschappen, waar computerondersteunde anamnese werd beproefd (zie o.a. Lucas, 1977). Nu kan gesteld worden dat het op dit moment niet zo'n probleem is om een standaardvragenlijst uit een enquête om te zetten in een interactief computerprogramma. Wat wél een probleem is, is of dat soort programma's ook door geënquêteerden zelfstandig gebruikt kunnen worden. Het gaat immers om een dialoogvorm die afwijkt van de gebruikelijke. Op deze kwaliteitsvraag én op de vraag of dataverzameling via de computer misschien meer mogelijkheden biedt dan alleen de vervanging van de ene enquêtevorm door de andere, gaan we in dit artikel wat nader in. In beide gevallen gaan we ervan uit dat een zogenaamde 'naïeve gebruiker' (iemand met geen of heel weinig ervaring met computers) direct converseert met de computer. Wij zien dus af van een behandeling van situaties waarin een enquêteur bijvoorbeeld via een telefoon vragen stelt aan een respondent en de antwoorden van de respondent invoert via een terminal (zie de Hond, 1981; Cannell, Groves, 1981).

2. ON-LINE DATAVERZAMELING

De karakteristieken van on-line dataverzameling vallen als volgt kort weer te geven:
- de respondent bevindt zich achter een standaard-display terminal met toetsenbord (we zien af van meer geavanceerde communicatiemedia als lichtpen en 'pad');
- vragen en antwoordmogelijkheden verschijnen op het scherm;
- de respondent kiest één van de antwoordmogelijkheden en tikt daarvan (meestal) het nummer in;
- de volgende vraag verschijnt, enzovoorts, totdat de enquête afgelopen is.

In concreto komen we dit in twee situaties tegen:
- de respondent komt naar een speciale experimenteerruimte;
- de 'interviewer' komt met een draagbare mini-computer bij de respondent thuis (zie Saris, de Pijper, Neijens, 1980).

Welke van deze twee situaties de voorkeur heeft, valt moeilijk te zeggen. De speciale experimenteerruimte heeft als voordeel dat het een rustige omgeving creëert, maar als nadeel dat de respondent moet reizen, hetgeen een grotere non-respons kan opleveren. Het thuis afnemen ondervangt het bezwaar van het reizen, maar eist bijzonder mobiele en betrouwbare apparatuur en legt bovendien een zwaar tijdsbeslag op instructeurs. Wat betreft voorkeuren van de respondenten zelf is er niet veel bekend. In een onderzoek (Bronner, De Hoog, 1982) stelden wij deze vraag aan een veertigtal proefpersonen. Daarvan gaf 20% de voorkeur aan het komen naar een experimenteerruimte (de situatie in dat onderzoek), terwijl 13% liever de enquêteur met mini-computer thuis had ontvangen. Tweederde gaf aan dat het hun niet veel uitmaakte.

Afgezien van de vraag welke situatie de voorkeur verdient, is natuurlijk ook

de vraag naar de voor- en nadelen van on-line dataverzameling zélf van belang. De voordelen zullen wij behandelen aan de hand van enkele kenmerkende onderdelen van het onderzoekproces.

Voordelen

Vragenlijstontwerp:
- complexe 'routing'-procedures kunnen opgenomen worden in het computerprogramma, waarna de respondent zich als het ware ongemerkt door de complexe lijst heenwerkt;
- rotatieprocedures, random presentatie, kunnen onder controle gehouden worden;
- de onderzoeker wordt gedwongen goed na te denken over de vragen die gesteld moeten worden. Immers, een 'gemakkelijke' open vraag kan meestal niet gesteld worden, omdat programma's de ingewikkelde antwoorden daarop (nog) niet kunnen verwerken;
- speciale vraagstellingstechnieken kunnen worden toegepast, technieken die via enquêteurs te grote moeilijkheden opleveren (daarop komen we straks nog terug);
- sommige nogal persoonlijke onderwerpen lenen zich beter voor enquêtering met de computer dan met een enquêteur (Lucas, et al., 1977).

Interviewer bias:
- het veldwerk kan op uniforme wijze en enquêteurs-onafhankelijk uitgevoerd worden (zie ook De Hond, 1981);
- interviewerfouten worden ondervangen, te denken valt onder andere aan het overslaan van vragen, het voorleggen van verkeerde antwoordalternatieven, het volgen van een verkeerde route, schrijf- en begripsfouten.

Coderen en ponsen:
- doordat de gegevens direct in geprecodeerde vorm worden opgeslagen, worden de tijdrovende en foutengevoelige fasen coderen en ponsen overgeslagen.

Analyse:
- het is in principe mogelijk om direct na het interview tot (standaard) analyse over te gaan. 'For example: regression analysis can be done and a plot showing the relationship between two variables can be made.' (Saris, et al. 1980).

Flexibiliteit:
- aanpassingen in de vragenlijst zijn, indien noodzakelijk, gemakkelijk en met geringe kosten uit te voeren gedurende de loop van het onderzoek.

Vatten wij deze voordelen samen, dan geldt dat er sprake is van een versnelling van het onderzoekproces, van flexibiliteit in de vragenlijstconstructie, van betere (meer precieze) data en van het stimuleren van het gebruik van nieuwe vraagstellingstechnieken.

Nadelen

- de methode is hoofdzakelijk bruikbaar wanneer het gaat om een onderwerp waarvan de vragen in gestructureerde vorm gesteld kunnen worden;
- wanneer de hardware en software niet *uiterst* betrouwbaar zijn, kan een sessie abrupt afgebroken worden, met alle vervelende gevolgen van dien;
- respondenten mogen weinig 'techniekangst' hebben en dienen enige typevaardigheid (zij het meestal in geringe mate) te bezitten. Uit door ons gehouden onderzoek (Bronner, De Hoog, 1982) viel op te maken dat de vereiste vaardigheid niet groot behoeft te zijn;
- omdat mensen vaak slecht lezen, dienen de vragen zo eenvoudig mogelijk te zijn. Complexe vragen met uitgebreide instructies en/of toelichtingen lijken niet aanbevelenswaardig, temeer daar ook een scherm maar een beperkte hoeveelheid informatie kan bevatten;
- omdat steun van de interviewer ontbreekt (uitleg, verheldering) moeten de vragen zo geformuleerd zijn dat ze in principe voor elke respondent in de steekproef begrijpelijk zijn; het is de vraag of dat wel te realiseren is;
- achteraf na enkele vragen een antwoord corrigeren is lastiger dan in een mondelinge gesprekssituatie;
- het toepassingsgebied lijkt voornamelijk nog beperkt te zijn tot kleine steekproeven. Het direct via de computer ondervragen van honderden respondenten, gespreid over het gehele land, is om praktische en financiële redenen momenteel nog onmogelijk. Dat zou kunnen veranderen wanneer informatiemedia als Viditel een bredere verspreiding onder de bevolking zouden vinden.

Samenvattend: de benadering legt beperkingen op aan het soort vragen dat gesteld kan worden, eist uiterst betrouwbare hard- en software, en is vooral bruikbaar bij kleinere steekproeven.

Uiteraard zit er aan deze nieuwe ontwikkeling ook een kostenaspect. Het is zonder meer duidelijk dat de procedure besparingen zal opleveren in de sfeer van codering en ponsen, deze worden in feite geheel overbodig. In hoeverre er ook besparing is in veldwerkkosten staat nog te bezien. Wanneer men kiest voor benadering van de mensen 'thuis' zullen er toch ook interviewers of instructeurs nodig zijn, terwijl er meer in apparatuur geïnvesteerd moet worden. Met een experimenteerruimte kan men op het aantal apparaten en interviewers besparen, maar dit brengt weer extra reiskosten van de respondenten met zich mee. Kosten die in ieder geval gemaakt zullen moeten worden zijn die van apparatuur, computertijd en programma-ontwerp en vooral het testen van programma's. Alles bij elkaar genomen zal er op korte termijn sprake zijn van een kostenverhoging. Of deze acceptabel is hangt af van verwachtingen op langere termijn en wenselijkheid van betere data.

3. UITBREIDING VAN MOGELIJKHEDEN

In het voorgaande hebben we ons vooral geconcentreerd op het enquêteren via de computer, waarbij deze laatste als het ware de taak van de interviewer overnam. Wanneer de mogelijkheden daartoe beperkt zouden blijven, dan zou men zich kunnen afvragen of tussenvormen zoals een telefonische enquête met

directe data-invoer door de interviewer niet zinniger en rendabeler zijn. Het aantrekkelijke van de computer is echter dat deze dataverzamelingstechnieken mogelijk maakt die in een gewone enquête vanwege de complexiteit niet of nauwelijks hanteerbaar zijn. Het gaat er dan niet meer om dat men iets per computer doet wat ook anders kan (vervanging), maar om iets dat eigenlijk alléén per computer kan (vernieuwing). De computer helpt dan bij het *uitbreiden* van het repertoire aan onderzoektechnieken. Aan dit aspect willen wij in deze paragraaf aandacht besteden. Eerst gaan wij in op een tweetal voorbeelden waarbij de computer al bestaande technieken nieuw leven inblaast, daarna vertellen we iets over een nieuwe techniek die voor een belangrijk deel via de computer is bedacht en ontwikkeld. De gepresenteerde technieken kunnen van belang zijn voor bijvoorbeeld produktbeoordeling en produktontwikkeling.

3.1. *Nieuwe kansen voor oude technieken*

Wij zullen ons beperken tot een tweetal voorbeelden: conjunct meten en de repertory grid.

Conjunct meten

Met behulp van de theorie van het conjunct meten kan men nagaan welke utiliteit of welk nut de consument toekent aan de verschillende waarden van attributen of eigenschappen van produkten (zie voor toepassing van deze techniek in het Nederlandse marktonderzoek van Raay, 1978 en Zeldenrust, 1979). Van Raay paste de techniek toe in een onderzoek naar gemalen koffie en komt tot conclusies zoals: 'een prijsverhoging van f 3,95 naar f 4,45 heeft een sterkere utiliteitsdaling tot gevolg dan een prijsverhoging van f 4,45 naar f 4,95. Een laag coffeïnegehalte heeft een groter utiliteitsverschil met een normaal coffeïnegehalte dan normaal met hoog'. De techniek wordt zeer belangrijk geacht om inzicht te krijgen in vaststelling van de trade-offs tussen attribuutniveau, vooral de prijs versus de overige produktattributen, maar blijkt in de praktijk in enquêtesituaties via enquêteurs moeilijk toepasbaar.

MacBride en Johnson (1980) en Blamires en Finer (1982) ontwikkelden een computerprogramma dat tot doel heeft die conjuncte meetprocedures makkelijker toepasbaar te maken. De kern is dat die procedures veel moeilijke, saaie en langdradige trade-offs vereisen. Via een interactief computerprogramma bleken respondenten deze trade-offs veel leuker en interessanter te vinden, waardoor veel meer mensen de procedure ook daadwerkelijk volbrachten en bereid waren beter over de trade-offs na te denken. Bovendien merken MacBride en Johnson op (blz. 51): 'Interview observations showed that interviewer's explanations of the conventional trade-off task were often inadequate. The electronic methodology is superior in this regard by virtue of being self-explanatory'.

Hieruit blijkt onder andere ook dat het eerder gesignaleerde nadeel van beperkte toelichtingsmogelijkheden in sommige gevallen ondervangen kan worden door ernaar te streven dat de vragen zo 'zelfuitleggend' mogelijk zijn.

Het kernvoordeel is echter dat deze trade-offs een *dynamische* structurering van het gesprek eisen, immers de volgende trade-off die aangeboden wordt

hangt af van het antwoord op de vorige vraag. Een dergelijke procedure is zonder computer in een 'gewone' enquêtesituatie vrijwel niet uit te voeren. MacBride en Johnson zeggen hierover (blz. 42): 'This process can be likened to playing chess with a computer. The computer continually evaluates new data (each trade-off response) and makes a new move (presenting the respondent with a challenging new trade-off decision to make)'.

De respondenten kregen 'frames' op het scherm te zien, waarvan voor verschillende soorten auto's een voorbeeld gegeven is (figuur 1).

De computerisering van de trade-off-techniek maakt het dus mogelijk om de theorie voor het conjunct meten in de toekomst op grotere schaal te gaan gebruiken. Daarmee zijn overigens ook nog niet alle problemen rond deze techniek opgelost.

Which would you prefer?

small size and medium speed	standard size and fast speed

(Remember these are the same in all other ways)

Strongly prefer left			No preference				Strongly prefer right	
1	2	3	4	5	6	7	8	9

Figuur 1. Voorbeeld van een trade-off zoals gebruikt in een computer-interactieve procedure (zie MacBride en Johnson, blz. 43).

Wanneer men met echt bestaande alternatieven werkt, zal het over het algemeen onmogelijk zijn om de positie van beide alternatieven in een trade-off op alle andere kenmerken constant te houden ('Remember these are the same in all other ways'). Dat kan alleen door te abstraheren van deze reële alternatieven, bijvoorbeeld door respondenten gelijkenissen te laten aangeven tussen die abstracties en de reële alternatieven (zie voor een voorbeeld Bronner, De Hoog, 1978). Ook daarbij kan de computer echter in dezelfde procedure mee behulpzaam zijn, doordat voor elk afzonderlijk individu de koppeling wordt bijgehouden tussen de geabstraheerde en reële alternatieven.

De repertory grid techniek

Een bekend probleem in het marktonderzoek is de vraag welke attributen, dimensies, eigenschappen, aspecten of kenmerken door mensen onderscheiden worden aan produkten of, meer algemeen, keuze-alternatieven.

Verschillende procedures zijn hiervoor ontwikkeld (verzamelen van gelijkenisgegevens gevolgd door meerdimensionale schaling, open vraagtechnieken, a priori gekozen attributen, semantische differentialen, repertory grid techniek).

Eerder werden door ons sterke en zwakke punten van deze procedures vergeleken (Bronner, De Hoog, 1979; Bronner, 1981). Ervaringen met de reper-

tory grid techniek leren dat de belangrijkste attributen naar voren komen en dat dit ook attributen zijn die daadwerkelijk in het proces van keuzebeïnvloeding te gebruiken zijn. De techniek is gebaseerd op Kelly's personal construct theory (Kelly, 1955).

Het idee in deze techniek is dat uit de verzameling keuze-alternatieven telkens verschillende drietallen worden geselecteerd en aangeboden aan de respondent. Deze wordt vervolgens gevraagd of hij/zij uit dit drietal een keuze-alternatief kan aanwijzen dat in enigerlei opzicht verschilt van de andere twee. Als het antwoord positief uitvalt wordt gevraagd of de respondent dit verschil in termen van twee polen of tegenstellingen kan uitdrukken. Een respondent geeft dus redenen voor onderscheiding, dit geeft de onderzoeker zicht op de gebruikte beoordelingsdimensies. Evenals bij conjunct meten geldt dat zeer bruikbare resultaten opgeleverd worden, maar dat toepassing in de praktijk veel problemen geeft. Sampson (1978, p. 33) zegt hierover: 'The information sought and collected is extremely specific. Although repertory grids may be administered, in conjunction with an individual intensive interview, if the grid is executed properly, it should be exhaustive and at the end of it, both the respondent and the interviewer will have had enough.'

Deze techniek leent zich echter ook uitstekend voor computerisering en in het boek van Shaw (1979) zijn een aantal computerprogramma's te vinden alsmede enkele uitbreidingen.

Wij hebben zelf ervaring opgedaan met de repertory grid in gecomputeriseerde vorm in het kader van een onderzoek naar een computerprogramma voor beslissingsondersteuning (zie De Hoog, Van Houten, 1980; Bronner, De Hoog, 1982). In dat programma zit de grid-techniek opgenomen als methode voor het eliciteren van relevante keuzekenmerken. Een voorbeeld van de door ons gehanteerde vraagstelling die op het scherm verscheen, staat in figuur 2.

Figuur 2. Voorbeeld van de grid-techniek zoals gebruikt in Bronner, De Hoog, 1982. (Antwoorden van respondenten in hokjes.)

Kunt u uit het volgende drietal partijen er een in contrast brengen tot de overige twee?
1) CPN
2) CDA
3) SGP
[...ja]

Welk nummer heeft de partij die zich van de beide andere onderscheidt?
[...1]
Waarin onderscheidt de CPN zich van de andere twee?
De CPN is nogal...
[...links]
CDA en SGP zijn daarentegen meer...
[...rechts]

Op basis van een eerste experiment, waarbij aan 40 proefpersonen gevraagd werd auto's en politieke partijen te beoordelen, is de formulering verbeterd. Een aantal duidelijke voordelen waren reeds te onderkennen: relatief veel ken-

merken worden genoemd, men wordt niet in de verleiding gebracht voorkeur-variante kenmerken zoals goed – slecht te noemen (zie over voorkeur-variantie Bronner, 1981) en het programma stelt mensen in staat hun eigen specifieke formulering voor een bepaald attribuut te geven. Zo bleek bijvoorbeeld dat er maar liefst 42 verschillende benamingen van het links/rechts-kenmerk in de politiek werden gehanteerd door 40 respondenten. Vooral omdat de proefpersonen hierna de partijen konden beoordelen op deze 'eigen' kenmerken, werd de 'persoonlijke' benaming als bijzonder prettig ervaren.

Hoewel dus de grid-techniek ook in principe 'met de hand' kan worden afgenomen, is deze voor de standaard-interviewsituatie toch vaak te complex. Het selecteren en aanbieden van drietallen, het registreren van antwoorden en benamingen van kenmerken, het controleren op voorkeur-variante kenmerken en het noemen van tweemaal hetzelfde kenmerk gaat allemaal zeer ordelijk en overzichtelijk via de computer en zonder registratiefouten. Het voordeel spreekt nog meer wanneer, zoals zo vaak in het marktonderzoek, ook een rating-procedure volgt. Dan kunnen via de computer direct de eigen karakteriseringen van de keuze-alternatieven aan de respondent voorgelegd worden, iets dat in een vragenlijst (vrijwel) onmogelijk is. Met andere woorden: de repertory grid, die wat in onbruik dreigde te geraken vanwege praktische problemen, kan via de computer-interactieve versie weer nieuw leven ingeblazen worden.

3.2. *Het maken van keuze-alternatieven*

Vaak gaat het er in het marktonderzoek niet alleen om de karakterisering van al bestaande alternatieven te achterhalen, ook het nog daarvóór liggende probleem, welke alternatieven zouden mensen eigenlijk willen hebben, staat in de belangstelling. In marktonderzoektermen: het gaat niet alleen om produktbeoordeling maar ook om produktontwikkeling. Computerprogramma's die helpen bij het maken van eigen keuze-alternatieven kunnen 'structuring-aid systems' genoemd worden. Recentelijk is door Pearl, Leal, Saleh (1980) het programma GODDESS ontworpen, waarmee iemand keuze-alternatieven kan ontdekken waarop deze anders niet gekomen zou zijn. Globaal werkt het programma als volgt:

- eerst wordt het hoofdprobleem/doel/wens vastgesteld;
- daarna komen de subdoelen aan de orde die bereikt/gerealiseerd moeten worden alvorens het hoofddoel aan bod komt;
- dan wordt gevraagd de actiestrategieën aan te geven die de subdoelen zouden kunnen realiseren;
- de uitwerking van de actiestrategieën wordt nagelopen;
- waarna de precondities daarvoor worden vastgelegd.

In figuur 3 staat een voorbeeld van deze uiteenrafeling in dialoogvorm en in figuur 4 een grafische weergave daarvan.

Figuur 3. Dialoog in een doelgericht zoekprogramma, zoals gebruikt door Pearl, et al., 1980, blz. 28 e.v.

Fasen	Vraag uit het programma	Voorbeeld van een antwoord
Goal	what is your problem?	improve my financial status
Subgoals	Let's analyze ways of achieving the main goal by considering its major objectives which must be achieved before the goal can be attained	have a high steady income
Action strategy	let's consider the objective 'have high steady income'; list all possible action strategies that you can take toward the fulfillment of 'have high steady income'; make sure that the list contains all physically feasible approaches (including possibly 'doing nothing') and that no two of them can be combined	1. open a business 2. find a second job 3. increase my overtime
Action mode	let's consider the action strategy 'open a business'; list the names for various detailed methods by which 'open a business' can be executed; idem 'find a second job'; idem 'increase my overtime'	1. franchise 2. joint-venture none none
Preconditions	let's consider the action mode 'open a business-franchise'; list any preconditions that must be established before mode 'open a business-franchise' can be implemented let's consider the action 'open a business-joint-venture'; list any preconditions that must be established before mode 'open a business joint-venture' can be implemented	none having management assistance
	etc. tot de boom resulteert die in figuur 4 is weergegeven.	

Ook het programma GODDESS wordt gekenmerkt door het feit dat de volgende actie in het programma afhangt van wat er in de vorige actie is geschied. Eveneens geldt dat het gebruiken van de eigen bewoordingen van het individu noodzakelijk is om het geheel zinvol te maken. Overigens zijn er ons op dit moment geen toepassingen van GODDESS in de sfeer van het marktonderzoek bekend. Wel lijken deze methoden de komende jaren een belangrijke rol bij produktontwikkeling te kunnen spelen.

Vatten wij deze paragraaf samen, dan illustreren de drie gegeven voorbeelden (conjunct meten, repertory grid, GODDESS) dat met behulp van de computer nieuwe of vernieuwde technieken in enquêtes geïntroduceerd kunnen worden. Wij zouden dit willen omschrijven als *dynamisch interviewen*. Dit dynamisch interviewen onderscheidt zich van het klassieke interview door het feit dat niet alleen qua routing verschillen kunnen optreden tussen individuen, maar ook en vooral doordat per individu de *inhoud* van de ene vraag afhangt van het

antwoord op een andere vraag (nogmaals: vergelijk dit met het spelen van schaak, waarbij de zet afhankelijk is van de laatste zet van de tegenstander). Daarnaast biedt het dynamisch interviewen per computer de mogelijkheid om 'persoonlijke' vragenlijsten te realiseren, dat wil zeggen vragenlijsten waarin de respondent herhaaldelijk zijn eigen formuleringen, zo men wil 'taal', zal tegenkomen in plaats van door de onderzoeker voor hem/haar uitgekozen woorden. Dat het ontwikkelen van een dergelijk hoogwaardig en respondentvriendelijk systeem speciale problemen oproept, willen we in de volgende paragraaf naar voren halen.

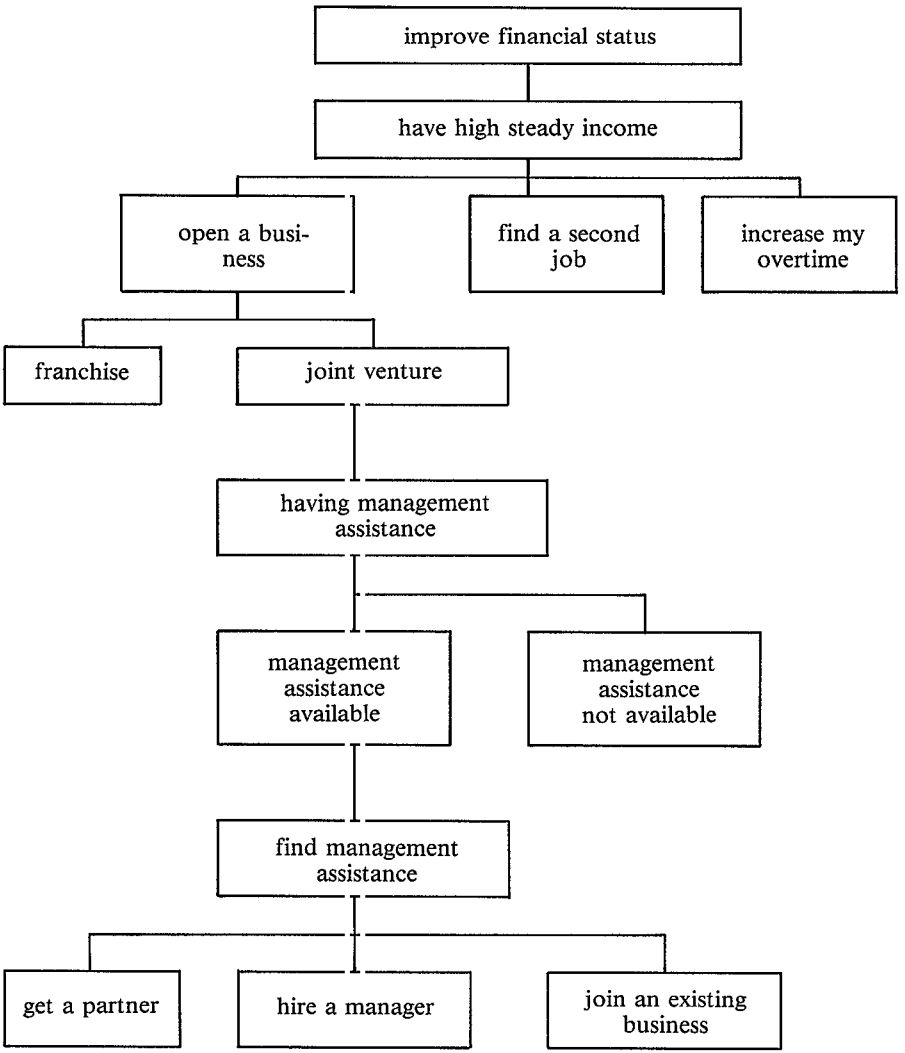
4. ONTWERP- EN ONTWIKKELINGSPROBLEMEN

Met computergestuurd interviewen zijn we natuurlijk ook meteen beland op het terrein van de ontwikkeling en kwaliteit van computerprogrammatuur. Een zeer uitgebreid terrein, met een enorme vracht aan onderzoek en literatuur die hier natuurlijk allemaal onmogelijk aan de orde gesteld kan worden. We willen ons daarom beperken tot een tweetal aspecten, die van bijzonder belang zijn wanneer men ertoe overgaat 'naïeve' respondenten of gebruikers direct in contact te brengen met de computer. Het gaat dan vooral om de *betrouwbaarheid* van de interactieve software en de *kwaliteit van de dialoog.*

4.1. Software-betrouwbaarheid

Het eerste en meest belangrijke punt is hier natuurlijk dat het computerprogramma als zodanig *correct* moet zijn, geen 'bugs' mag bevatten. Niets is vervelender voor iemand aan een terminal dan dat het programma er opeens om onverklaarbare reden mee stopt, meestal gevolgd door een reeks onbegrijpelijke foutmeldingen. Het programma moet dus eerst uitgebreid getest worden. Het is noodzakelijk alle mogelijke 'routes' minstens één keer doorlopen te hebben. Dat kan bij een wat complexer enquêteprogramma nog heel wat tijd en moeite kosten.

Ten tweede moet rekening gehouden worden met onverwachte acties van de kant van de naïeve gebruiker. Wanneer iemand bijvoorbeeld een positie op een rating-schaal moet aangeven tussen 1 en 9 en per ongeluk wordt 99 ingetikt, dan moet het programma daarvan niet in de war raken of gewoon doorgaan met fouten verderop. Het proces van het ondervangen van deze onverwachte acties wordt ook wel 'bullet proofing' (Martin, 1973) genoemd. De kern van dit proces wordt gevormd door de eis dat het programma voor *elke* gebruikersactie een zinvolle respons heeft. Deze respons kan de vorm hebben van het wijzen op fouten met het verzoek ze te corrigeren tot een algemene hulpfaciliteit die op elk moment aangeroepen kan worden. Het zorgen voor dit type faciliteiten kan veel inspanningen vragen. Het zal in de praktijk vaak zo zijn dat een 'bullet proof'-programma voor het grotere deel bestaat uit onderdelen die onverwachte gebruikersacties moeten opvangen, terwijl het 'eigenlijke' programma veel kleiner is. Een voorbeeld van zo een onverwachte gebruikersactie kunnen wij ontlenen aan ons eigen experiment. Op bepaalde vragen konden de proefpersonen 'ja' of 'nee' antwoorden en elk antwoord dat niet 'nee' luidde, werd als 'ja' geïnterpreteerd. Wij hadden er geen reke-



Figuur 4. Grafische representatie van een probleemstructuur zoals die uit de dialoog met de computer naar voren komt (Pearl et al., 1980, blz. 44).

ning mee gehouden dat sommige respondenten het antwoord 'neen' zouden intikken, hetgeen door het programma als 'ja' geïnterpreteerd werd.

De investeringen hierin moeten dus niet onderschat worden. Martin (1973) bevat een aantal procedures die kunnen helpen bij het totstandbrengen van een goed interactief systeem.

4.2. *Kwaliteit van de dialoog*

De dialoog tussen respondent en ondervragende computer gaat uiteraard in de vorm van taal. Daarbij komen twee typen problemen naar voren. Ten eerste

lezen de meeste mensen tamelijk slecht. Ten tweede hoeft taal die voor de één begrijpelijk is, dat niet noodzakelijkerwijs ook voor de ander te zijn. Beide factoren maken dat bij het ontwerpen van een computergestuurde enquête de grootst mogelijke zorgvuldigheid in acht genomen moet worden. Nu is het in sociaal-wetenschappelijk onderzoek gebruikelijk om veel aandacht te besteden aan de formulering van enquêtevragen, zodat deze expertise al goed gebruikt kan worden. In het dynamische interview met een computer komt daar echter bij dat ook de taal van de geënquêteerde zelf gebruikt kan worden. Zolang computers nog maar op zeer beperkte schaal gewone, dagelijkse ('natuurlijke') taalinputs kunnen verwerken, zal de dialoog in dit geval vaak een 'template mode' aannemen. Daaronder wordt verstaan dat in de tekst van het programma ruimte wordt opengelaten, en in de daadwerkelijke communicatie met de gebruiker de relevante woorden van de gebruiker zullen worden ingevuld (zie bijvoorbeeld de conversatie in figuur 3). Omdat nog niet bekend is wat elke afzonderlijke gebruiker zal gaan zeggen, moet zorgvuldig aandacht worden besteed aan de omringende tekst, zodat geen al te onverwachte taal- of zinsconstructies optreden. Die kunnen de serieusheid waarmee de gebruiker met het programma communiceert, ondergraven.

Uiteraard is dit maar een klein onderdeel van het totale dialoogprobleem. De ruimte ontbreekt echter om daar dieper op in te gaan. De geïnteresseerde lezer wordt verwezen naar Martin (1973) en voor een overzicht naar Shneiderman (1980, hoofdstuk 11) en Smith, Green (1980).

5. TOT SLOT

In dit artikel hebben we aandacht besteed aan de mogelijkheden voor het gebruik van computers in enquêtesituaties. Wij hebben ons beperkt tot situaties waarbij de respondent *zelf* direct converseert met de computer. Geconstateerd is dat het in principe niet zo moeilijk is om een standaardenquête per computer uit te voeren. Gezien echter de kosten en de beperkingen – qua omvang – van de steekproef, zal toepassing op deze wijze voorlopig beperkt blijven. Op korte termijn lijkt het computer-interactief enquêteren vooral goede mogelijkheden te bieden als het gaat om het nieuw leven inblazen van complexe maar adequate dataverzamelingsmethoden. Aan de hand van een drietal voorbeelden hebben we duidelijk gemaakt dat het kan gaan om een *dynamisch interview,* waarbinnen oude en nieuwe onderzoektechnieken ingezet kunnen worden ter verhoging van de kwaliteit van de gegevens. De ontwikkeling van dit soort computerprogramma's eist echter dat er uiterste zorg wordt besteed aan het ontwerpen van een goede dialoog en het creëren van een 'bullet-proof'-systeem. Wanneer aan die voorwaarden is voldaan, wordt het verstandig en lonend om de computer in te schakelen bij enquêteringsactiviteiten in het marktonderzoek.

Van een vervanging van de enquêteur zal op korte termijn geen sprake zijn, het zal veeleer gaan om een aanvulling. Met name zal toepassing plaats gaan vinden in het kleinschalige onderzoek.

BIBLIOGRAFIE

Blamires, C., P. Finer, 1982. *Micros and modelling: putting a new slant on established techniques,* Proceedings Market Research Society Conference, Brighton, pp. 277-289.

Bronner, A.E., 1981. *Ontwikkelingen in image-onderzoek,* Jaarboek van de Ned. Ver. v. Marktonderzoekers, pp. 3-18.

Bronner, A.E., R. de Hoog, 1978. *Politieke voorkeur: oordelen en beslissen,* diss. Universiteit van Amsterdam.

Bronner, A.E., R. de Hoog, 1979. *Keuzemodellen en marktonderzoek,* Jaarboek v.d. Ned. Ver. van Marktonderzoekers, pp. 223-239.

Bronner, A.E., R. de Hoog, 1982. *Non-expert use of a computerized decision aid,* in P.C. Humphreys, O. Svenson, A. Vari (eds.) Analysing and Aiding Decision Processes, North Holland.

Cannell, C.F., R.M. Groves, 1981. *Comparison of results between computer assisted telephone interviewing (CATI) and paper and pencil questionnaires,* SRC, University of Michigan.

Hond, M. de, august 1981. *New international research technique 'rings the bell' throughout Europe,* Proceedings 34th ESOMAR-Congress, Amsterdam, pp. 219-245.

Hoog, R. de, H.J. van Houten, 9.5. 1980, *Keuze-agogie per computer,* Tijdschrift voor Agologie, pp. 381-404.

Kelly, G.A., 1955. *Psychology of personal constructs, vols. I, II,* New York.

Lucas, R.W., 1977. *Questioning patients by computer,* Health Bulletin, pp. 296-302.

Lucas, R.W., P.J. Mullini, C.B. Luna, D.C. McInroy, 1977. *Interrogation of patients by computer,* British Journal of Psychiatry, pp. 131, 160.

MacBride, J.N., R.M. Johnson, september 1980. *Respondent reaction to computer-interactive interviewing techniques,* Proceedings 33rd ESOMAR-Congress, Monte Carlo, pp. 39-54.

Martin, J., 1973. *Design of Man-Computer Dialogues,* Prentice Hall, New Jersey.

Pearl, J., A. Leal, J. Saleh, 1980. *GODDESS: a goal-directed decision structuring system,* University of California, L.A.

Raay, W.F. van, 1978. *Het gebruik van de conjuncte meettechniek,* Jaarboek van de Ned. Ver. van Marktonderzoekers, pp. 187-208.

Sampson, P., 1978. *Qualitative research and motivation research,* in: R. Worcester, J. Downham (eds.): Consumer market research handbook, New York.

Saris, W.E., W.M. de Pijper, P. Neijens, 1980. *Some notes on the computer steered interview,* Dutch Sociometric Society.

Shaw, M.L.G., 1979. *On becoming a personal scientist,* Academic Press.

Shneiderman, B., 1980. *Software psychology,* Winthrop.

Smith, H.T., T.R.G. Green, 1980. *Human Interaction with Computers,* Academic Press, London.

Zeldenrust, D., 1979. *De prijs als attribuut – een voorbeeld van conjunct meten en simulatie,* Jaarboek van de Ned. Ver. van Marktonderzoekers, pp. 135-146.

3. Het gebruik van panelgegevens t.b.v. sales decision support-systemen

A.J. OLIVIER

1. INLEIDING

Een groot aantal bedrijven is in het bezit van panelgegevens op basis waarvan men de eigen positie op de markt op continue basis in beeld wil brengen.

In de Nederlandse situatie heeft men de keuze uit een retailpanel, dat door Nielsen is opgezet en het consumentenpanel, dat door Attwood in stand wordt gehouden.

Een verkoper zal zich primair baseren op de ex factory-omzetten. Dit omdat de door het bedrijf geformuleerde verkoopstrategie uiteindelijk resulteert in de maand- of weektarget die in tonnen, geld of eenheden wordt geformuleerd.

Ook de bijbehorende klantenstrategieën worden op basis van de eerder genoemde grootheden geformuleerd.

Voor een goede analyse van de verschillen tussen de doelstellingen en de werkelijke uitkomsten is er echter behoefte aan aanvullende informatie.

Wij denken hierbij aan informatie over:
- de totaalmarkt
- de omzet van de concurrentie
- de omzetontwikkeling bij klanten
- de omzetontwikkeling per regio
- de op de markt gehanteerde prijzen
- de verkrijgbaarheid van de produkten
- de ontwikkelingen van de diverse marktsegmenten.

Het is haast ondenkbaar vandaag de dag een goed verkoopbeleid te voeren zonder deze ondersteunende externe achtergrondinformatie. Vele fabrikanten en importeurs zijn derhalve in het bezit van een panelservice.

Niet alle gebruikers zijn op de hoogte van de tekortkomingen en eigenaardigheden van deze panel-data en vaak worden de door het marktonderzoekbureau geleverde cijfers ten onrechte als een volkomen perfecte weergave van de werkelijkheid gezien.

Verkoopfunctionarissen die deze gedachte accepteren en hun eigen ex factory-gegevens zonder meer vergelijken met de panelcijfers zullen van de ene verbazing in de andere vallen en de verkoopfunctionarissen die deze vergelijking niet maken, lopen het risico acties te ondernemen die uitsluitend gebaseerd zijn op misverstanden.

In deze paper wordt nader ingegaan op de vergelijking tussen ex factory-gegevens en panelgegevens, waarbij de districtvergelijkingen centraal staan.

Aangetoond zal worden dat een vergelijking zonder meer niet mogelijk is, terwijl daarnaast aangegeven zal worden op welke wijze men door een integratie van beide cijferbronnen de kwaliteit van het panel als tool of management kan verbeteren.

2. DEKKING VAN PANELGEGEVENS

Indien panelgegevens vergeleken worden met werkelijke ex factory-gegevens dan stuit men onmiddellijk op het begrip dekking. Onder dekking wordt dan verstaan: de mate waarin de hoeveelheden produkt die de fabriek verlaten, wordt teruggevonden door een retail panel of een consumentenpanel. Deze dekking is afhankelijk van een aantal factoren.

1. Weglekken van de omzet naar andere kanalen.
2. Oneigenlijke exporten/importen van produkt.
3. Voorraadvorming.
4. Non-response-vertekeningen.

Ad 1 *Weglekken van de omzet naar andere kanalen*

Een retail panel heeft een strakke omschrijving van de kanalen waarover gerapporteerd wordt.

Gerenommeerde lekken worden gevonden in de Horeca-sector, amusements- en sportsector en de alom voortschrijdende parallellisatietendensen.

Ad 2 *Illegale importen en exporten*

In de meeste verkooporganisaties zijn binnenlandse en buitenlandse verkopen strikt gescheiden zaken.

Op een groot aantal markten worden sommige produkten uit een markt in verhoogde mate door buitenlanders aangekocht. Daarnaast geldt voor importeurs nog wel eens het probleem dat deze produkten afkomstig van een buitenlandse collega, onverwachts (vaak tegen dumpprijzen) op de binnenlandse markt opduiken.

Ad 3 *Voorraadvorming*

De groothandel en de detailhandel hebben handelsvoorraden die in de tijd gezien in omvang kunnen toe- of afnemen en daar meer verschillen kunnen doen ontstaan tussen af-fabriek verkopen en door het panel gemeten omzet.

Ad 4 *Non-response-vertekeningen*

Ook wanneer de drie hierboven genoemde factoren geen rol spelen, mag niet verwacht worden dat de geproduceerde en verkochte omzet perfect geregistreerd worden.

De bureaus beheersen weliswaar de techniek om steekproeffluctuaties tot een minimum te beperken, maar hebben geen verweer tegen die respondenten die weigeren mee te werken aan het onderzoek.

Hoewel het onderzoekbureau dit rechttrekt door een andere vergelijkbare respondent te ondervragen, slaagt men er niet volledig in systematische vertekeningen te voorkomen. Zo is een aantal goedkope merken en kanalen binnen een consumentenpanel overdekt, terwijl bij een retail panel juist een aantal dure merken overdekt is.

3. DE DEKKING VAN PANELGEGEVENS PER GEWEST

Voor een fabrikant van fast moving consumer goods voor de Nederlandse levensmiddelenmarkt werden ex factory-gegevens uit 1980 vergeleken met panelgegevens, waarbij de omzet voor een volledig jaar in beeld gebracht werd teneinde voorraadfluctuaties zo goed mogelijk te elimineren.

Tabel 1.

Dekking per Nielsen-district	Retail panel	Consumer panel
Nationaal	78	72
Nielsen I + II	88	83
Nielsen III + IV	79	74
Nielsen V	68	61

Conclusies

- Zowel door het retail panel als door het consumer panel worden de omzetten redelijk goed gedekt.
- De dekkingen per Nielsen-gebied lopen wat verder uiteen. Ongeveer 20%.
- Dit verschijnsel is zowel bij een consumer panel als bij een retail panel aanwezig en vertoont dezelfde trend.

Wanneer de regionale verfijningen nog wat verder gevoerd worden tot het niveau van verkoopdistricten, dan beginnen de dekkingen nog wat verder uit elkaar te lopen. Hierbij dient dan te worden aangetekend dat een consumer panel en een retail panel niet gestratificeerd worden op de verkoopgewesten van een fabrikant en dat de statistische onbetrouwbaarheid derhalve per definitie aanmerkelijk toeneemt.

Bij een fabrikant van fast moving consumer goods werd een vergelijking gemaakt tussen zijn afleveringen per verkoopdistrict en de omzetten zoals deze door de panels geregistreerd worden. Omdat deze indeling gedeeltelijk parallel loopt met de Nederlandse provincies, spreken wij in het vervolg over provincies. Onderstaande tabel toont de resultaten:

Conclusies

- De provincies Zeeland, Limburg, Drente en Overijssel zijn overdekt. Dit zou betekenen dat er meer produkt geconsumeerd wordt dan er wordt afgeleverd.
- Utrecht en Noord-Brabant zijn onderdekt. Dit betekent dat er gedeelten van de ex factory-omzetten niet worden teruggevonden.

4. OORZAKEN VAN DE DEKKINGSVERSCHILLEN

Alvorens in te gaan op de gevolgen van deze vertekeningen voor verkoop is het noodzakelijk in te gaan op de oorzaken van de overdekking per district. Deze oorzaken moeten toegevoegd worden aan de verschijnselen die in paragraaf twee behandeld werden en die in hun algemeenheid altijd van toepassing zijn.

Tabel 2. Dekking per provincie

Dekking per verkoopgewest	Retail panel	Consumer panel	Verschil
Friesland/Groningen	67	75	−8
Drente/Overijssel	115	98	17
Gelderland	93	81	12
Utrecht	60	54	6
Noord-Holland	83	77	6
Zuid-Holland	69	63	6
Zeeland	120	120	0
Noord-Brabant	57	53	4
Limburg	104	86	18

4.1 *De toerekening van ex factory-omzetten*

Het feit dat ex factory-gegevens afleveringen zijn op groothandelsniveau en panelgegevens op detaillistenniveau, is een belangrijke oorzaak van de dekkingsverschillen. De klanten van een groothandel zijn bijna altijd over meerdere provincies verspreid.

Grote G.W.B.- en V.F.B.-organisaties kennen één of een aantal centrale afleveringspunten. De mate van dekking is derhalve afhankelijk van de mate waarin de centrale afleveringspunten in een verkoopgewest aanwezig zijn. Binnen een rationeel opgebouwde fysieke distributiestructuur kan er, door uit te gaan van paneldekkingen in een land als Nederland, een aantal goederenstromen getraceerd worden. Een goederenstroom kent als startpunt een gebied met een lage dekking (er is produkt verdwenen) en kent als eindpunt een hoge dekking (er is meer geconsumeerd dan er afgeleverd is).

Op basis hiervan kunnen de volgende goederenstromen getraceerd worden.

Tabel 3. Traceren van fysieke goederenstromen op basis van dekkingsverschillen

Van	Naar
Utrecht/Zuid-Holland	Noord-Holland
Zuid-Holland/Noord-Brabant	Zeeland
Noord-Brabant	Limburg
Utrecht	Gelderland/Overijssel
Groningen/Friesland	Overijssel/Drente

Hierbij zijn ongetwijfeld alleen de hoofdstromen op deze markt in beeld gebracht. De werkelijkheid zal complexer zijn en minder eenduidig. Alleen de netto-effecten werden derhalve in beeld gebracht.

Toch illustreert dit voorbeeld duidelijk hoe groot de verschillen zijn tussen ex factory-cijfers en panelcijfers. Deze dekkingsfactoren gelden voor het totale assortiment van de fabrikant. Omdat in een district aan meerdere organisaties afgeleverd wordt en deze organisaties allemaal een ander assortiment artikelen van de fabrikant afnemen, kan de dekking per keer zeer verschillen en ook in de tijd meer afwijken.

4.2. *De vergelijking tussen plaats van aankoop en plaats van consumptie*

Een tweede potentiële storingsmogelijkheid is het feit dat de consument niet in het verkoopgewest koopt waarin zij woont.

Het feit dat de retail-paneldekkingen (kopen) en de consumer-paneldekkingen (consumeren) erg dicht bij elkaar liggen, bewijst dat de dekkingsverschillen ontstaan in de fase van groothandel naar detaillist en niet in de fase van detaillist naar consument.

De geringere mobiliteit van de consument ten opzichte van de handelaar heeft tot gevolg dat het hierboven beschreven type aankoop zonder twijfel aanwezig is, maar zich statistisch uitmiddelt en hierdoor een factor van geringe betekenis wordt.

De substantiële verschillen die waarneembaar zijn tussen retail-dekkingen en consumer-dekkingen zijn niet het gevolg van goederenstromen, maar worden veroorzaakt door de grenshandel. Vandaar dat in de districten Limburg, Gelderland en Overijssel de grootste verschillen worden aangetroffen tussen ex factory-omzetten en panelomzetten. In het verkoopgewest Friesland/Groningen vindt nauwelijks grenshandel plaats. Het grote verschil tussen retail-panel- en consumer-panel-uitkomsten wordt hier juist veroorzaakt door de hoge non-respons-vertekening.

4.3 *Statistische onbetrouwbaarheid*

Daarnaast wordt het steekproefmateriaal bij een uitsplitsing van de panelgegevens naar zoveel regionale gebieden wat onbetrouwbaarder dan de nationale cijfers.

Vooral kleine merken met lage spreidingen verliezen snel aan nauwkeurigheid.

5. DE BETROUWBAARHEID VAN DE RECHTSTREEKSE LEVERINGEN

Een gedeelte van de omzet wordt meestal rechtstreeks aan het filiaal geleverd en op basis van dit gegeven is het onderscheid tussen de fouten, die afkomstig zijn van dekkingsverschillen en de afwijkingen die ontstaan ten gevolge van de fysieke goederenstroom, nauwkeurig in kaart te brengen.

In principe moeten de goederen die rechtstreeks aan een retail outlet geleverd worden, ook zo goed als volledig door consumenten uit die provincie geconsumeerd worden.

Tabel 4.

Dekking per verkoopgewest	Rechtstreekse leveringen	Groothandels-leveringen	Totaal
Nationaal	82	69	72
Nielsen I + II	85	71	74
Nielsen III + IV	92	81	83
Nielsen V	74	55	61
Friesland/Groningen	101	71	75
Drente/Overijssel	101	87	98
Gelderland	89	78	81
Utrecht	94	45	54
Noord-Holland	71	80	77
Zuid-Holland	83	59	63
Zeeland	101	128	120
Noord-Brabant	61	51	53
Limburg	80	92	86

Aan deze tabel kunnen de volgende conclusies ontleend worden:

- De rechtstreekse leveringen worden beter gedekt dan de groothandelsafleveringen.

  Vooral het nationale cijfer en het cijfer per Nielsen-district leveren een bevredigend beeld op. Alleen Nielsen V, waarin een belangrijk stuk grenshandel zit, levert een afwijking op.
- Ook het beeld per provincie wordt aanmerkelijk verbeterd door deze uitsplitsing.

  In een groot aantal districten wordt 90-100% van de rechtstreekse leveringen door het consumentenpanel getraceerd. Dat dit in Noord-Brabant en Limburg niet het geval is, moet geweten worden aan de grenshandel. In Zuid- en Noord-Holland vallen de dekkingen ook tegen, hetgeen toegeschreven moet worden aan consumentenaankopen door inwoners van de andere districten, of het feit dat gedeelten van de rechtstreekse leveringen toch naar andere verkoopgewesten getransporteerd worden.

6. DE VERGELIJKING VAN EX FACTORY-ONTWIKKELINGEN MET DE PANELONTWIKKELINGEN

In de vorige hoofdstukken werd geïnventariseerd door welke oorzaken ex factory-cijfers en panelcijfers van elkaar kunnen afwijken.

Deze afwijkingen kunnen voor het merendeel verklaard worden, waarbij het verschil tussen plaats van aflevering en plaats van consumptie de belangrijkste oorzaak bleek.

Het feit dat de rechtstreekse leveringen aan detailhandelsfilialen aanmerkelijk beter gedekt worden is hiervan het formele bewijs.

De uitsplitsing naar verkoopgewest kan derhalve zinvolle beleidsinformatie opleveren indien deze informatie met kennis van zaken geïnterpreteerd wordt. Alvorens in het volgende hoofdstuk een voorstel uit te werken tot een integratie van ex factory-cijfers en panelcijfers zal in dit hoofdstuk geïnventariseerd

worden wat er gebeurt wanneer deze integratie niet plaatsvindt.

Allereerst een theoretische uitwerking van de informatie die de verkoopfunctionaris op dit moment tot zijn beschikking heeft. Deze informatie kan betrekking hebben op:

- een klant
- een verkoopgewest
- een organisatie
- een outlettype.

Aan de hand van deze informatie kan de verkoopfunctionaris zijn target vergelijken met de door hem gerealiseerde omzet. Uitgewerkt is de situatie waarin de verkoopmanager in september 1981 terugblikt op het eerste halve jaar 1981.

Tabel 5. Organisatie X

eenheden	jan. t/m juni 1980	jan. t/m juni 1981	index actual	nationale index	target index
merk 1	50	55	110	95	105
merk 2	30	25	83	83	105
merk 3	10	12	120	105	105
merk 4	5	10	200	130	150
totaal	95	102	107	94	107

In dit vereenvoudigde voorbeeld is goed te zien hoe men een organisatie (klant of gewest) in beeld kan brengen. Indexen t.o.v. de vergelijkbare periode vorig jaar laten zien welk merk in welke mate gegroeid of gedaald is.

Deze ontwikkeling kan vergeleken worden met het landelijk beeld, waaruit blijkt dat deze organisatie met haar index van 107 voor ligt op het landelijk beeld.

Dit wordt in dit voorbeeld dan weer voor een belangrijk gedeelte verklaard door de groei die voor deze organisatie gepland is: 5% voor de bestaande produkten.

Merk 4 is net geïntroduceerd en kent derhalve een aparte target van 150.

Een aantal belangrijke vragen wordt door dit type overzichten echter niet beantwoord:

- is onze omzet meegegroeid met de organisatie;
- welke indexen worden door de concurrent bereikt.

Dit type overzichten staan verkoopmanagers ten dienste als decision support-systeem bij de uitvoering van de door de fabrikant gevoerde geweststrategie.

Ook de gewest/district-managers hebben dit type overzichten die het gebied vanuit hun specifieke optiek in beeld brengen.

Voor een gewestelijk verkoopmanager van een fabrikant van fast moving goods hebben wij de conclusie van zijn ex factory vergeleken met de uitsplitsing binnen het retail panel.

Het gewest van deze verkoopmanager bestaat uit vier verkoopdistricten.

Het gaat hier om een vergelijking tussen ex factory-cijfers en consumentenpanelcijfers.

Tabel 6. Overzicht van de significante verschuivingen binnen het gewest van de verkoopmanager

	Gewest	District 1	District 2	District 3	District 4
merk 1	R	ER	E	ER	–
merk 2	E	R	E	ER	E
merk 3	–	ER	E	–	ER
merk 4	E	–	–	ER	E
merk 5	E	–	–	ER	ER
merk 6	R	ER	R	ER	–
merk 7	E	E	E	ER	ER
merk 8	E	E	E	E	E
merk 9	–	ER	E	E	–
merk 10	R	E	E	R	ER

E = significante beweging in de ex factory-cijfers.
R = significante beweging in de consumer panel-cijfers.

In deze rumoerige en turbulente markt waren de meeste merken (80%) in beweging, hetzij de ex factory-cijfers, hetzij de cijfers, zoals die door het consumer panel gemeten worden. Slechts in 35% van de gevallen waarin een significante verandering plaatsvindt is er sprake van een situatie waarin ex factory-cijfers en consumer panel-cijfers elkaar niet tegenspreken. Het is duidelijk dat er voor een verkoopmanager die met deze discrepanties tussen de cijferbronnen geconfronteerd wordt, weinig zinvols te analyseren valt.

Zij zijn in ieder geval niet zonder meer met elkaar vergelijkbaar. Men kan zich afvragen of er niet een integratie van beide cijferbronnen denkbaar is, waardoor een voor de verkoopmanager praktisch hanteerbaar tool of management bestaat. Hierover gaat het volgende hoofdstuk.

7. DE INTEGRATIE VAN EX FACTORY DATA EN PANEL DATA TOT EEN HANTEERBAAR DECISION SUPPORT-SYSTEEM VOOR DE DIVERSE VERKOOPDISCIPLINES

De kritische kanttekeningen uit de vorige hoofdstukken hadden geenszins de intentie om een van beide databronnen in een kwaad daglicht te stellen, maar om vanuit de tekortkomingen van de databronnen een bruikbaar geïntegreerd alternatief te ontwikkelen die de tekortkomingen van beide databronnen omzeilt.

Hierbij zijn de volgende uitgangspunten gehanteerd:

- verkoopmanagers mogen nooit geconfronteerd worden met elkaar tegensprekende cijfers.

  In het meest extreme geval zou men alleen de meest betrouwbare informatie moeten krijgen en zou de daarmee in strijd zijnde informatie niet verstrekt moeten worden.
- verkoopmanagers krijgen expliciet de cijfers met betrekking tot de taakvelden waarvoor zij zelf verantwoordelijk zijn. Rechtstreekse leveringen aan individuele outlets en aan afleveringspunten van grootwinkelbedrijven dienen derhalve buiten de overzichten gehouden te worden.

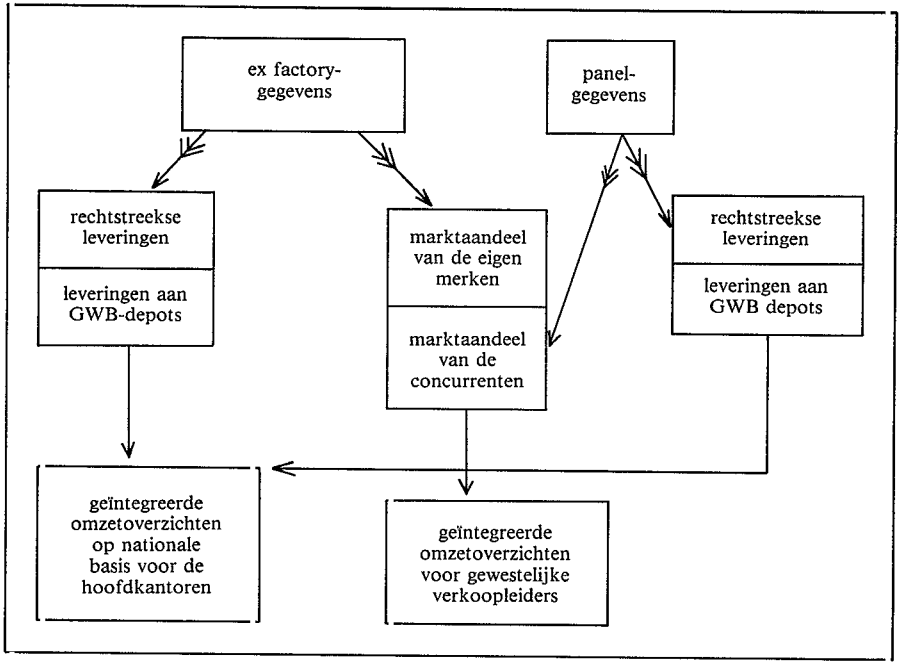
  Dit laatste impliceert dat ex factory-cijfers centraal zullen blijven staan.

– verkoopmanagers zullen zo veel mogelijk cijfers moeten krijgen die hen in staat stellen oorzaak-en-gevolg-relaties te leggen bij het interpreteren van de verkoopcijfers.

Dit impliceert dat ook informatie m.b.t. de concurrenten in een sales decision support system verwerkt zou moeten worden.

Deze concurrentiegegevens zijn vanzelfsprekend niet op ex factory-niveau beschikbaar en zullen uit een andere bron verkregen moeten worden.

Uit meerdere cijferbronnen zal derhalve een geïntegreerd nieuw bestand gegevens geconstrueerd moeten worden. Dit kan als volgt schematisch in beeld gebracht worden.



Schema 1. Omzetoverzichten per verkoopdiscipline

Hoe de daadwerkelijke integratie het best tot stand kan komen kan het beste geïllustreerd worden aan de hand van een concreet cijferbeeld.

Hierbij wordt voortgebouwd op het cijfervoorbeeld uit tabel 5.

De gegevens binnen het kader zijn cijfers die afkomstig zijn uit de originele databestanden. De andere gegevens zijn hieruit afgeleid.

De panelcijfers dekken de totale afzet maar gedeeltelijk. In 1980 is dit niveau 75% (DEF). In 1981 loopt dit niveau licht terug.

De dekking per merk verschilt in de tijd aanmerkelijk. Sommige dekkingen verslechteren, terwijl andere verbeteren. Een essentieel gegeven is het feit dat de richting waarin de omzetten per week zich bewegen, gelijk is.

Wij gaan uit van de veronderstelling dat de mate waarin onze produkten in

hun totaliteit gedekt worden, gemiddeld genomen ook van toepassing is voor de concurrenten.

$$\text{D.w.z.: } PF/EFF = PT/EFT\,. \tag{7.1}$$

Ten gevolge van deze veronderstelling zal ook het marktaandeel binnen de ex factory-cijfers per definitie gelijk zijn aan het marktaandeel dat gevonden wordt op basis van de panelgegevens.

$$\text{D.w.z.: } EFF/EFT = PF/PT\,. \tag{7.2}$$

De totaalmarkt binnen het desbetreffende district kan nu als volgt ingeschakeld worden.

$$EFT = PT \times EFF/PF\,. \tag{7.3}$$

Terwijl voor de individuele concurrent geldt dat:

$$C_x = EFT \times PC_x/PT\,. \tag{7.4}$$

Dezelfde exercitie kan nu ook voor 1981 uitgevoerd worden. De resultaten van deze berekeningen zijn opgenomen in tabel 7.

Op basis van deze berekeningen kunnen de groei-indexen uit tabel 5 uitgebreid worden met de groei-indexen van de concurrerende fabrikanten.

Deze indexen kunnen naar analogie van tabel 5, waarin de werkelijke indexen afgezet worden tegen de nationale index en de target index, vergeleken

Tabel 7. District X

basis: eenheden per week	omzet per week ex factory		omzet per week-panel		dekking %		index
	jan. t/m juni 80	jan. t/m juni 81	jan. t/m juni 80	jan. t/m juni 81	jan. t/m juni 80	jan. t/m juni 81	
merk 1	50	55	40	44	80	80	110
merk 2	30	25	18	12	60	48	83
merk 3	10	12	10	14	100	116	120
merk 4	5	10	3	5	60	50	200
subtotaal fabrikant	95 (EFF)	102	71 (PF)	75	75 (DEF)	73	107
concurrent 5	34 (C5)	35	25 (PC5)	26	73	74	103
concurrent 6	27 (C6)	41	20 (PC6)	25	74	73	126
concurrent 7	46 (C7)	41	35 (PC7)	30	76	73	89
concurrent 8	9 (C8)	13	7 (PC8)	10	77	77	144
totaal 1	211 (EFT)	225	158 (PT)	166	75	74	107

Verklaring van de tekens:
EFF = ex factory – omzet van de fabrikant.
PF = panel – omzet van de fabrikant.
DEF = ex factory-dekking. D.w.z. het deel van de feitelijke hoeveelheid produkt dat de fabriekspoort verlaten heeft en dat uiteindelijk door een panel teruggevonden wordt wanneer de steekproefuitkomsten worden opgeblazen naar een landelijk niveau.
EFT = totaalmarkt op ex factory-basis.
PT = totaalmarkt op panel-basis.
PCX = panel-omzetten van concurrent.
CX = ex factory-omzet van concurrent.

worden met onze schattingen van de concurrentie-ontwikkelingen en het landelijk beeld.

Op basis van deze tabel kan nu een nieuwe definitieve tabel geconstrueerd worden waarin alle panelgegevens zijn weggelaten en waarin opgenomen:
- gem. omzetten per week
- marktaandelen
- groei-indexen
- landelijke indexen
- target/prognose indexen.

Conclusies

De informatie die verkoopmanagers ter beschikking gesteld kan worden blijkt te kunnen worden verbeterd door het toevoegen van retail panel-informatie of consumentenpanel-informatie aan de veelal bestaande overzichten op basis van ex factory-gegevens.

Het zonder meer ter beschikking stellen van deze informatie zal echter leiden tot een hoge mate van onzekerheid bij de verkoopmanagers als gevolg van de tegenstrijdige conclusies die uit de diverse cijferbronnen getrokken kunnen worden. Deze verschillen zijn voor een belangrijk deel te verklaren vanuit het feit dat beide cijferbronnen op een andere plaats de goederenstroom meten en derhalve per definitie andere resultaten geven.

Door het beter vergelijkbaar maken van de gemeten goederenstroom en het integreren van beide cijferbronnen ontstaat een eenduidig beeld.

Een belangrijk gevolg van deze integratie is het feit dat de omzetten van de concurrent de verkoopmanager in staat stellen beter vast te stellen welke rol de concurrenten in het gebied (of de organisatie) waarvoor hij verantwoordelijk is, gespeeld hebben.

De absolute hoogte van het marktaandeel van de concurrent kan weliswaar door dekkingsverschillen vertekend zijn, maar de trend zal er altijd voldoende duidelijk in naar voren komen.

Wel zal het beschikbaar stellen van geïntegreerd cijfermateriaal altijd kritisch en professioneel begeleid moeten worden.

BIBLIOGRAFIE

Leeflang, P.S.H. and Reuyl J.C., 'De bepaling van de effectiviteit van marketing instrumenten'. *Tijdschrift voor Marketing* jrg. 16, nr. 6, pp. 3-10.

Leeflang, P.S.H. and Olivier A.J., 'Market Intelligence monitoren, het beeld v/d jaren '80', *Tijdschrift voor marketing,* jrg. 15, nr. 6, pp. 10-15, 1981.

Leeflang, P.S.H. and Duyn, J.J. van, 'On the use of regional in marketing models, The Demand for beer in the Netherlands, Part II'. *European Research,* Vol. 10, 1982.

Leeflang, P.S.H. and Olivier, A.J., 'Facing panel non-response, consequences and solutions'. *Proceedings 35rd Esomar Congress,* 1982, pp. 417-440.

Leeflang, P.S.H. and Olivier, A.J., 'What is wrong with the audit data we use for decision making in marketing', *proceedings 33rd Esomar Congress,* 1980 pp. 219-239.

Mensing, W.C.J.J, The representation of non cooperating chains in the Nielsen retail index, *Unpublished Nielsen paper* 1981.

Zwan, A. van de, 'Grondslagen en techniek van de marktanalyse', 1980, pp. 179-181.

4. Foutenbronnen in steekproefonderzoek*

H.M.P. KERSTEN EN J.G. BETHLEHEM

SAMENVATTING

De kwaliteit van de uitkomsten van steekproefonderzoek wordt bepaald door de steekproeffout en de niet-steekproeffout. De relatie tussen deze twee fouten wordt aangegeven en een uitsplitsing van de niet-steekproeffout in verschillende soorten wordt gegeven. De foutenbronnen kunnen in verschillende fasen van het steekproefonderzoek optreden: bij het opstellen van de vragenlijst, bij de vaststelling van de doelpopulatie, de keuze van de trekkingseenheid, bij het steekproefkader en bij het trekken van de steekproef. Bij het veldwerk kunnen de meeste foutenbronnen tegelijk optreden. Elk van de fasen van het steekproefonderzoek wordt besproken en voorbeelden worden gegeven.

1. INLEIDING

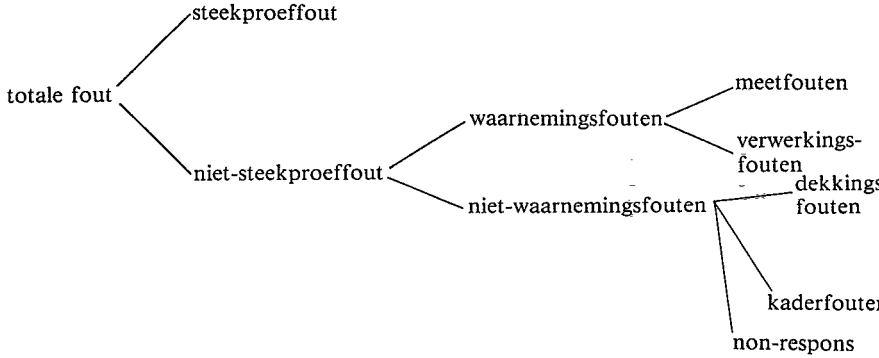
Het kwaliteitsaspect van steekproefonderzoek krijgt de laatste jaren terecht meer aandacht. Op verschillende plaatsen (waaronder het CBS) wordt hiernaar onderzoek verricht. Uit de statistische consultatie blijkt dat er zich een groot aantal problemen kunnen voordoen bij het opzetten, uitvoeren, bewerken en verwerken van een steekproefonderzoek. Door de ontwikkelingen op het gebied van de automatisering kunnen een aantal foutenbronnen (ponsfouten, coderingsfouten, inconsistenties e.d.) onder controle worden gehouden, waardoor de aandacht gericht kan worden op andere verschijnselen die invloed uitoefenen op de uitkomsten van steekproefonderzoek. In dit artikel wordt ingegaan op de verschillende soorten fouten die mogelijk zijn, de fase(n) van het onderzoek waarin men er hinder van kan ondervinden en wat de consequenties daarvan kunnen zijn. Hoewel er in dit artikel steeds gesproken wordt over fouten, impliceert dit niet altijd dat er van foutieve handelingen sprake is, maar wordt het woord fout ook gebruikt om onzekerheid aan te duiden. De relatie tussen de steekproeffout en de niet-steekproeffout wordt besproken en een nadere differentiatie van de niet-steekproeffout wordt gegeven. Door meer aandacht te schenken aan de aard en het effect van deze fouten kan de kwaliteit van het onderzoek worden verbeterd. Het onderscheiden van een aantal fasen bij het steekproefonderzoek en daarin optredende foutenbronnen maakt het mogelijk specifieke problemen te onderkennen. Aan de hand van voorbeelden uit de praktijk wordt de invloed van de fouten geïllustreerd en worden recente publicaties op het desbetreffende onderzoeksgebied vermeld.

* De meningen weergegeven in dit artikel zijn die van de auteurs en komen niet noodzakelijk overeen met die van het CBS

2. DE TOTALE FOUT

Steekproefonderzoek is één van de manieren om meer informatie over een te onderzoeken populatie te verkrijgen. Het kenmerkende van steekproefonderzoek is, dat op grond van gegevens omtrent een gedeelte van de populatie uitspraken worden gedaan over de gehele populatie. Wil steekproefonderzoek zin hebben dan moeten er procedures beschikbaar zijn om enerzijds steekproeven op correcte wijze te trekken, en anderzijds om op grond van de steekproefgegevens uitspraken over de populatie te doen. Aangezien nooit absoluut zekere uitspraken kunnen worden gedaan, is het van belang om de onzekerheid in de uitkomsten te kwantificeren.

Bij het typeren van de mogelijke fouten maken we gebruik van een indeling zoals die gegeven is door Bethlehem en Kersten (1982a). Dit schema vindt zijn oorsprong in Kish (1967) en wordt in figuur 1 gegeven.



Figuur 1. Indeling van de totale fout in steekproefonderzoek

Allereerst wordt de totale fout gesplitst in een steekproeffout en een niet-steekproeffout. De *steekproeffout* ontstaat doordat men aan slechts een gedeelte van de elementen in de populatie waarnemingen verricht. Op grond van de beschikbare waarnemingen wil men door middel van een *schatting* een uitspraak doen over de waarde van een populatieparameter. Het recept om een dergelijke schatting te berekenen wordt een *schatter* genoemd. Bij herhaald (onder gelijkblijvende omstandigheden) uitvoeren van de steekproeftrekking zal de schatting niet steeds dezelfde waarde hebben. De afwijking tussen de met behulp van de schatter in de steekproef gevonden waarde en de werkelijke waarde wordt de steekproeffout genoemd. In het algemeen kan de onderzoeker door een geschikte keuze van het steekproefontwerp en de te gebruiken schatter, de steekproeffout onder controle houden en kwantificeren. Er zijn echter situaties denkbaar waarbij de elementen andere trekkingskansen hebben dan de onderzoeker bedoeld heeft. De oorzaak hiervan kan liggen bij de nog te behandelen dekkingsfouten en kaderfouten. Fouten die hierdoor ontstaan vallen vaak niet in de praktijk te ontdekken.

Steekproeffouten treden niet op wanneer aan elk element in de populatie metingen worden verricht. We spreken dan van een *integrale telling* of *census*. Toch hoeft de waarde van de populatieparameter die op grond van een integrale telling berekend is, niet overeen te komen met de werkelijke waarde. De

fouten die in deze situatie gemaakt worden, noemen we de *niet-steekproeffouten*. De niet-steekproeffouten kunnen door een veelheid van factoren worden veroorzaakt. Voor een nadere bestudering van deze fouten is het zinvol om deze fout te splitsen in componenten.

Daartoe wordt de niet-steekproeffout allereerst gesplitst in waarnemingsfouten en niet-waarnemingsfouten. Zoals de naam al aangeeft ontstaan *waarnemingsfouten* door fouten die worden gemaakt in het veld bij het verzamelen en vastleggen van de gegevens. De waarnemingsfouten die in het veld worden gemaakt noemen we *meetfouten*. Ze ontstaan doordat kennelijk de gevraagde gegevens op onjuiste wijze verstrekt worden. Enerzijds kan de oorzaak liggen bij de enquêteur/trice doordat hij/zij de vragen niet juist stelt, de ondervraagde beïnvloedt, of de antwoorden onjuist registreert, anderzijds kan de fout ook liggen bij de ondervraagde, doordat hij de vragen niet begrijpt of bewust verkeerde antwoorden geeft. Ook de vragenlijst zelf kan een bron van fouten zijn. In het verwerkingsproces dat plaatsvindt na de gegevensverzameling en waarin de gegevens gecontroleerd en vastgelegd worden, kan eveneens het nodige misgaan. *Verwerkingsfouten* kunnen ontstaan door het verkeerd coderen van antwoorden, het maken van ponsfouten, etc.

Naast fouten als gevolg van verkeerde waarneming en/of registratie kunnen ook fouten ontstaan doordat de gewenste metingen in het geheel niet verricht kunnen worden. Men noemt de fouten die hierdoor ontstaan *niet-waarnemingsfouten*. Voor het trekken van een steekproef moet men in het algemeen de beschikking hebben over een *steekproefkader* (een lijst van namen of adressen) en een methode om uit dit kader een steekproef te trekken. Indien dit steekproefkader niet exact samenvalt met de doelpopulatie van het onderzoek kunnen problemen ontstaan. Is dit het geval dan spreken we van *dekkingsfouten* (Engels: 'coverage errors'). Bevat het steekproefkader elementen die niet tot de doelpopulatie behoren dan spreken we van *overdekking* (Engels: 'overcoverage'). Zo zal men bij een bejaardenonderzoek niet de beschikking hebben over een steekproefkader waarin uitsluitend bejaarden zijn opgenomen en via welke alle bejaarden in Nederland kunnen worden bereikt. Gebruik van het persoonsregister voor dit onderzoek zal veel overdekking opleveren. De oplossing voor dit probleem is eenvoudig: geselecteerde elementen, die niet tot de doelpopulatie behoren, worden genegeerd en daarvoor in de plaats wordt nogmaals getrokken. Grotere problemen doen zich voor in het geval dat er elementen zijn die wèl tot de doelpopulatie behoren maar niet in het steekproefkader zijn opgenomen. Dit probleem van *onderdekking* (Engels: 'undercoverage') doet zich bijvoorbeeld voor ten aanzien van illegaal in ons land verblijvende personen (indien een trekking uit een persoonsregister plaatsvindt) en bij adressensteekproeven waarbij eerst een inventarisatie wordt gemaakt van de personen die op dat adres wonen. Indien de student op kamers niet wordt opgegeven bij de inventarisatie is er sprake van onderdekking. Het is immers onmogelijk om hem via het kader van adressen te bereiken. Niet-waarnemingsfouten kunnen ook optreden indien het desbetreffende element wel gelokaliseerd wordt, maar geen gegevens worden verkregen. Dan spreken we van *non-respons*. De oorzaak hiervan kan enerzijds liggen bij de te ondervragen persoon doordat hij niet in staat is of niet bereid is om te antwoorden (niet-thuis, weigering, ziekte), anderzijds bij de enquêteur/trice

doordat hij/zij de desbetreffende persoon niet bezoekt (adres bestaat niet, huis onvindbaar, huis afgebroken, gevaarlijke hond, etc.). Naast de problemen van onder- en overdekking speelt ook de kwaliteit van het steekproefkader een rol. *Kaderfouten,* dat wil zeggen vermeldingen van foutieve gegevens of het meerdere malen (ten onrechte) opgenomen zijn van elementen in het kader, hebben hun weerslag op het al dan niet in de steekproef trekken van elementen, met alle gevolgen van dien. Problemen met het steekproefkader werken door in de trekkingskansen van de elementen en beïnvloeden aldus mede de steekproeffout. Elementen die meerdere keren in het kader zijn opgenomen hebben een grotere trekkingskans dan elementen die hierin slechts éénmaal opgenomen zijn.

De hierboven beschreven terminologie is in de meeste steekproefonderzoeken toepasbaar, maar meer ingewikkelde onderzoeken kunnen problemen geven. Wanneer het steekproeftrekken in verschillende fasen plaatsvindt, bijvoorbeeld bij een meertrapssteekproef, kan de terminologie voor elke fase worden gehanteerd. Een foutenbron kan in de ene fase een andere classificatie krijgen dan in een andere fase. Wanneer bijvoorbeeld eerst een steekproef van adressen wordt getrokken en met één persoon per adres een vraaggesprek wordt gehouden, dan is het verzwijgen van de student op kamers zowel een meetfout in de eerste fase als een (onder)dekkingsfout in de tweede fase. Men kan het als een meetfout beschouwen aangezien de inventarisatie op het adres niet correct is. Er is sprake van onderdekking in de fase van enquêtering, aangezien de kans van de student om in de steekproef te worden getrokken gelijk is aan nul.

Een wat andere indeling van de totale fout, die enigszins los staat van de indeling hiervoor, kan worden gemaakt door te kijken naar de componenten waaruit de fout is opgebouwd. We onderscheiden twee componenten: de variantie en de vertekening. De *variantie* meet de ruis om de gemiddelde waarde van de schatter bij het onder dezelfde omstandigheden herhalen van het onderzoek. De ruis veroorzaakt fouten in de schattingen die niet systematisch te hoog of te laag zijn. De *vertekening* meet de systematische afwijking tussen de schatting en de werkelijke waarde bij het onder dezelfde omstandigheden uitvoeren van het onderzoek. Dekkingsfouten zullen in het algemeen aanleiding geven tot een vertekening. Indien bij de steekproeftrekking geen selectiefouten optreden, heeft bij de meeste situaties de steekproeffout alleen een variantiecomponent. Waarnemingsfouten kunnen leiden tot zowel een vertekenings- als een variantiecomponent.

3. WANNEER TREDEN WELKE FOUTEN OP?

3.1. *Inleiding*

Om na te gaan in welke fasen van het steekproefonderzoek foutenbronnen kunnen optreden, zullen we bij de opzet en uitvoering van het onderzoek een aantal fasen onderscheiden en per onderdeel aandacht besteden aan problemen die kunnen voorkomen. Het blijkt dat meerdere foutenbronnen tegelijk kunnen optreden. Dit wordt geïllustreerd aan de hand van voorbeelden uit de praktijk. Waar mogelijk zal ook worden ingegaan op correctieprocedures.

De verschillende onderdelen hangen zeer nauw met elkaar samen en bij de opzet van het steekproefonderzoek wordt in principe met alle onderdelen van het onderzoek rekening gehouden. Zo is het bijvoorbeeld ondenkbaar dat bij het steekproefontwerp geen rekening wordt gehouden met het steekproefkader en de ophoging. We behandelen achtereenvolgens de logische stappen die moeten worden gezet vanaf het moment dat bijvoorbeeld een opdrachtgever iets wil (laten) onderzoeken tot de uiteindelijke publicatie over dat onderzoek. Nadat het doel van het onderzoek is vastgesteld en de populatie waarover uitspraken worden verlangd omschreven is, komen steekproefaspecten aan de orde. Het steekproefontwerp moet worden gemaakt, waarna de steekproef kan worden getrokken en het veldwerk plaatsvindt. (Soms wordt de steekproef tijdens het veldwerk getrokken, men denke aan het trekken van een steekproef van adressen waarbij per getrokken adres aselect een persoon wordt aangewezen die gevraagd wordt aan het onderzoek deel te nemen.) In de veldwerkfase kunnen veel fouten optreden. Bij de ophoging (dit is het maken van totaalschattingen, op grond van de gegevens van de respondenten waarbij eventueel gecorrigeerd wordt voor effecten van non-respons en dekkingsfouten, voor de gehele doelpopulatie) wordt geprobeerd de nauwkeurigheid van de uitkomsten te vergroten, hetgeen betekent dat de precisie vergroot en/of de vertekening verkleind wordt.

3.2. *De formulering van de doelstelling van het onderzoek*

Deze fase, waarin de opdrachtgever duidelijk maakt om welke doelstelling (variabelen, samenhang e.d.) het gaat, is van zeer groot belang voor het slagen van het onderzoek. Wanneer er vaagheid bestaat omtrent het gebruik dat de opdrachtgever wil maken van de door het onderzoek te leveren uitkomsten, kan dit in een latere fase grote problemen oproepen en worden verwachtingen niet beantwoord. Een onderscheid dat van wezenlijk belang is in deze fase, is of het de bedoeling van het onderzoek is om *relaties* te onderzoeken of dat de doelstelling is het verkrijgen van *totaalcijfers of gemiddelden* voor de gehele doelpopulatie.

Voorbeelden van onderzoek waarbij relaties centraal staan (ook wel *relationeel steekproefonderzoek* genoemd) zijn het Nationaal kiezersonderzoek (o.a. wat verklaart het stemgedrag), het Leefsituatie-onderzoek en het onderzoek Inkomenswaardering (wat beïnvloedt de waardering van het inkomen). Steekproefonderzoeken waarbij het schatten van (nationale) totalen centraal staat (ook wel *niveau-steekproefonderzoek* genoemd) zijn bijvoorbeeld het Woningbehoeftenonderzoek (o.a. het schatten van het aantal woningzoekenden) en de Arbeidskrachtentelling (o.a. het schatten van het aantal werkzoekenden, forensen e.d.). De scheiding tussen relationeel steekproefonderzoek en niveau-steekproefonderzoek is overigens niet strikt.

Er is een duidelijke relatie tussen de doelstelling van het steekproefonderzoek en het steekproefontwerp dat ervoor wordt gehanteerd. Onder het steekproefontwerp verstaat men de beschrijving van het kansmechanisme waarmee de in de steekproef te trekken elementen kunnen worden verkregen. Een goed steekproefontwerp voor een onderzoek dat totaalschattingen moet opleveren, is een ontwerp dat de totale fout, behorende bij een schatting van een totaal,

minimaliseert. Aangezien relationeel steekproefonderzoek een heel andere doelstelling heeft, zijn voor dit soort onderzoek de schattingen van de verschillende correlatie- en regressiecoëfficiënten van veel meer belang en dient de totale fout hierin te worden geminimaliseerd. Dit kan tot een heel ander steekproefontwerp leiden. Een bijzonder voorbeeld van relationeel steekproefonderzoek wordt beschreven door Aigner (1979). De enige doelstelling van dit onderzoek was het schatten van een regressievergelijking tussen twee variabelen. In dat geval moet het steekproefontwerp leiden tot een minimale fout in de geschatte regressiecoëfficiënten. Hij laat de eventuele vertekening buiten beschouwing, hetgeen betekent dat het steekproefontwerp moet leiden tot een minimale variantie behorende bij de geschatte regressiecoëfficiënten. Met behulp van de formules waarmee deze varianties kunnen worden geschat, kan men de optimale allocatie van de in de steekproef te trekken elementen bepalen.

3.3. *De vaststelling van de doelpopulatie*

De bevolkingsgroep waarover uitspraken worden verlangd, wordt meestal bepaald door de doelstelling van het onderzoek. In het algemeen hoeft deze doelpopulatie niet exact samen te vallen met de gehele populatie. Voorbeelden van een doelpopulatie zijn: personen van 18 jaar of ouder, alleenstaanden, huisvrouwen, ziekenhuispatiënten, enz. De doelpopulatie hoeft ook niet exact samen te vallen met de steekproefkaderpopulatie, noch met de populatie van potentiële respondenten, ook wel het responsstratum van de populatie genoemd. Murthy (1979) onderscheidt nog een vijfde populatie, namelijk de populatie waar naartoe wordt opgehoogd. Deze populatie kan verschillen van de gehele populatie, de doelpopulatie en de populatie van elementen die opgenomen zijn in het steekproefkader. Men kan zich voorstellen dat indien het aantal woningzoekenden in Nederland moet worden geschat, het tijdsaspect een rol kan spelen. De steekproef van het Woningbehoeftenonderzoek 1981 is bijvoorbeeld in de zomer van 1981 getrokken en de personen die in de steekproef zijn opgenomen worden aan het eind van het jaar geënquêteerd. De aldus verkregen antwoorden hebben betrekking op de bevolking zoals die bestond gedurende de zomer, maar beschrijven bijvoorbeeld de huisvestingssituatie zoals die tijdens het najaar was. We zullen dit verschil tussen doelpopulatie en ophoogpopulatie hier verder niet beschouwen, hoewel er in de praktijk regelmatig rekening mee moet worden gehouden. In deelparagraaf 3.5 wordt dieper ingegaan op de relatie die er bestaat tussen de doelpopulatie, het steekproefkader en de respondenten.

3.4. *Het steekproefontwerp*

Onder het steekproefontwerp (Engels: 'sample design') verstaat men de beschrijving van de manier waarop de personen die in het veld zullen worden ondervraagd, in de steekproef worden getrokken. Een aspect dat hierbij een rol speelt is het verschil dat er bestaat tussen de trekkingseenheid en de waarnemingseenheid. We beperken ons in deze deelparagraaf tot twee steekproeven: personensteekproeven en adressensteekproeven. Bij de eerste categorie wordt

een persoon in de steekproef getrokken en alleen deze persoon wordt ondervraagd, terwijl bij de tweede categorie *alle* personen op het in de steekproef getrokken adres worden ondervraagd. De methode waarbij aan één persoon gevraagd wordt om ook voor alle andere personen op het adres te antwoorden (men denke aan het bezit van bepaalde goederen) kan worden ingedeeld bij de adressensteekproeven. Men kan zich voorstellen dat het enquêteren van meerdere personen op één adres over persoonsgebonden kenmerken soms minder informatie geeft dan het enquêteren van het zelfde aantal personen over deze kenmerken op verschillende adressen. Een maat die aangeeft hoe efficiënt gemeten (in aantal ondervraagden) het is om alle personen op het adres te ondervragen is het zogenaamde methode-effect (Engels: 'design effect'). Het methode-effect is hier gelijk aan het quotiënt van de variantie, behorende bij de uitkomsten wanneer alle personen op het adres worden ondervraagd en de variantie, behorende bij de uitkomsten indien slechts één persoon per adres wordt ondervraagd (bij een gelijk aantal ingevulde enquêteformulieren). De grootte van het effect hangt nauw samen met de variabelen die moeten worden onderzocht.

Tabel 1. Het methode-effect voor enkele kenmerken voor vier gemeenten

Kenmerken	Amsterdam	Rotterdam	Marum	Schinnen
Man	0,69	0,56	0,53	0,54
Ongehuwd	1,02	1,00	0,89	0,89
Gehuwd	1,12	1,04	0,89	0,92
Gescheiden	1,19	1,09	1,22	1,11
Weduwstaat	1,10	1,15	1,11	1,04
Jonger dan 15 jaar	1,51	1,33	1,41	1,28
65 jaar en ouder	1,72	1,72	1,80	1,55
Autochtoon	2,16	2,10	1,88	2,42
Rooms-Katholiek	3,09	3,20	4,01	2,58
Nederlands-Hervormd	2,42	2,63	3,99	2,38
Gereformeerd	2,97	3,28	4,48	1,00
Andere godsdienst	3,52	2,82	5,99	3,28
Geen godsdienst	2,86	2,80	3,51	2,42
Lage opleiding	1,46	1,59	1,58	1,48
Hoge opleiding	1,22	1,37	1,32	1,29

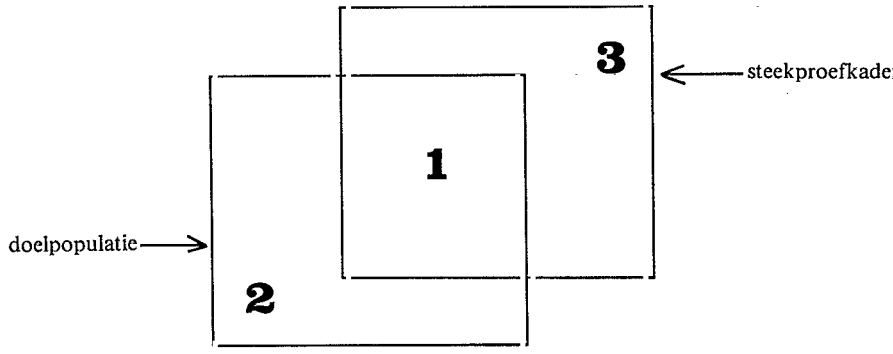
Tabel 1, die ontleend is aan De Ree en Nieuwenbroek (1981), geeft het methode-effect voor een aantal kenmerken voor de gemeenten Amsterdam, Rotterdam, Marum en Schinnen. De resultaten zijn gebaseerd op gegevens van de Volkstelling 1971. Uit deze tabel kan men bijvoorbeeld afleiden dat indien men het aantal personen in Amsterdam dat Rooms-Katholiek is wil schatten (of variabelen wil onderzoeken die hiermee sterk samenhangen), het aantal personen dat men moet ondervragen wanneer men alle personen op één adres enquêteert, driemaal zo groot moet zijn als wanneer men slechts één persoon per adres enquêteert om een zelfde precisie (= variantie) te bereiken. Uit de tabel blijkt duidelijk dat de grootte van het methode-effect niet voor verschillende variabelen en gemeenten gelijk is. De resultaten zoals gegeven in De Ree en Nieuwenbroek (1981) bieden de mogelijkheid om een geschikte keuze te maken tussen het enquêteren van één persoon per adres en het enquêteren

van alle personen op het adres. In de praktijk wordt de keuze tussen deze mogelijkheden mede bepaald door de haalbaarheid (veldwerk, steekproefkader e.d.). De kennis van het methode-effect kan dan toch nog worden gebruikt bij het berekenen van de variantie behorende bij de uitkomsten. Voor wat betreft het kostenaspect van de enquêtes geldt dat (zoals in bovenstaand voorbeeld) een verdrievoudiging van het aantal enquêtes niet leidt tot een verdrievoudiging van de kosten, aangezien de meerkosten van een enquête op een adres waarop reeds iemand ondervraagd is, geringer zijn dan de kosten van één enquête op een adres.

Het is eenvoudig mogelijk om van waarnemingseenheid over te gaan op trekkingseenheid of omgekeerd. We moeten dan wel de beschikking hebben over gegevens die het mogelijk maken om de kans van de eenheid waarover uitspraken worden gedaan, te berekenen. Bij wijze van voorbeeld kan het Woningbehoeftenonderzoek worden genoemd, waarbij gewerkt wordt met een personensteekproef. De persoon is de trekkingseenheid. Wanneer echter uitspraken moeten worden gedaan over huishoudens of woningen geldt dat de trekkingskans van woningen waarin veel personen wonen, groter is dan de trekkingskans van andere woningen. Wanneer echter bekend is uit hoeveel personen het huishouden bestaat of hoeveel personen er in de woning wonen, kan de trekkingskans herberekend worden. De methode die soms door marktonderzoekbureaus wordt gebruikt, waarbij aselect één persoon wordt aangewezen op een in de steekproef getrokken adres, valt onder de categorie van personensteekproeven. Ook in dit geval moet strikt genomen rekening worden gehouden met verschillen in trekkingskans van personen in grote of in kleine gezinnen.

3.5. *Het steekproefkader*

Veel problemen kunnen bij het veldwerk worden voorkomen indien er gewerkt kan worden met een goed steekproefkader. Het steekproefkader kan worden omschreven als een verzameling elementen waardoor elementen van de doelpopulatie kunnen worden bereikt. Het steekproefkader hoeft niet exact samen te vallen met de doelpopulatie. Men spreekt dan van *dekkingsfouten*. In figuur 2 wordt een schematisch overzicht gegeven van de relatie tussen de doelpopulatie en het steekproefkader.



Figuur 2. Schematische voorstelling van de doelpopulatie en het steekproefkader

Er kunnen zich verschillende mogelijkheden voordoen. We zullen ze hier puntsgewijs bespreken.

1. Hiermee komen elementen overeen die zowel tot de doelpopulatie behoren als via het steekproefkader kunnen worden bereikt.
2. Hiermee komen elementen overeen die weliswaar tot de doelpopulatie behoren maar niet via het steekproefkader kunnen worden bereikt. Deze elementen kunnen niet benaderd worden. Men spreekt hier van *onderdekking*. Voorbeelden hiervan zijn illegale gastarbeiders (indien de steekproef uit bevolkingsregisters wordt getrokken) en de rijdende en varende bevolking (indien de steekproef uit een adressenbestand wordt getrokken).
3. Hier bevinden zich elementen die weliswaar in het steekproefkader zijn opgenomen maar niet tot de doelpopulatie behoren. Men spreekt van *overdekking*. Voorbeelden hiervan zijn kinderen (indien het onderzoek alleen personen van 18 jaar of ouder betreft) en niet-autobezitters (indien het gaat om een onderzoek naar de rijsnelheid op autosnelwegen).

Het schema in figuur 2 heeft vooral betrekking op de situatie waarbij er een steekproef uit een echt steekproefkader wordt getrokken, maar kan ook worden gehanteerd indien de steekproeftrekking in het veld plaatsvindt (het werken met een aselect startadres of het aselect selecteren van een persoon op een adres).

Naast dekkingsproblemen is het ook mogelijk dat er fouten in het kader zelf zitten. Men spreekt van *kaderfouten* indien de elementen van de doelpopulatie weliswaar in het kader zitten maar niet correct, bijvoorbeeld met een foutief adres of met een verkeerde leeftijd. Deze fouten worden meestal veroorzaakt door het feit dat de administratieve weergave van de doelpopulatie wat achterloopt bij de werkelijke doelpopulatie. Een indicatie voor de kwaliteit van het steekproefkader kan worden verkregen door de resultaten van het veldwerk (men denke aan redenen van non-respons bij een personensteekproef als 'persoon verhuisd', 'huis staat leeg', 'huis afgebroken' en 'persoon overleden') te relateren aan de gegevens als vermeld in het steekproefkader.

De ordening of de opbouw van het steekproefkader is van belang. Het kan een alfabetische ordening zijn, maar er zijn ook steekproefkaders voor personensteekproeven waarbij de elementen zijn gerangschikt naar adres, wijk of leeftijd. Met een dergelijke ordening moet rekening worden gehouden bij het vaststellen van regels die moeten worden gehanteerd bij het trekken van de steekproef. Soms bevat het steekproefkader aanduidingen die het mogelijk maken om een gestratificeerde steekproef te trekken. Deze mogelijkheid doet zich bijvoorbeeld voor bij het afgiftepuntenbestand van de PTT, waarbij een aantal perceelscodes kunnen worden onderscheiden. Een voordeel van het afgiftepuntenbestand is dat het bijgewerkt is en dat geautomatiseerd trekken van de steekproef mogelijk is. Door het CBS is o.a. in het kader van de Arbeidskrachtentelling 1981 onderzoek verricht naar de kwaliteit en volledigheid van het afgiftepuntenbestand. Bij het Consumenten Conjunctuur Onderzoek wordt aan de enquêtrices gevraagd om extra gegevens over het in de steekproef getrokken adres te verschaffen zoals bijvoorbeeld: hoeveel huisnummers zijn er op het adres, hoeveel huishoudens zijn er in werkelijkheid, enzovoorts. Eveneens worden de hardheid en de inhoud van de perceelscode onderzocht. Eén resultaat van dit onderzoek is dat wellicht 25% van de afgifte-

punten met perceelscode 0 (dat wil zeggen overige percelen: verpleegstershuizen, cafés e.d.), die meestal door marktonderzoekbureaus buiten het onderzoek worden gelaten, ook bewoond worden.

Een van de grootste problemen die zich kunnen voordoen is het ten onrechte dubbel voorkomen van elementen in het steekproefkader. Deze elementen hebben dan in vergelijking met andere elementen een tweemaal zo grote kans om te worden getrokken. (Bijvoorbeeld: een winkelier die naast de winkel woont kan op beide adressen geënquêteerd worden.) Aangezien meestal niet van elk element bekend is of het dubbel voorkomt of niet, kan hiervoor niet of nauwelijks gecorrigeerd worden.

Tot slot zij opgemerkt dat voor specifieke doelpopulaties zoals bijvoorbeeld artsen, advocaten, agrariërs, enzovoorts, specifieke steekproefkaders kunnen bestaan. Door gebruik te maken van dergelijke kaders kan men veel tijd en kosten besparen, aangezien het probleem van overdekking geringer is.

3.6. *Het trekken van de steekproef*

3.6.1. *Inleiding*

Het hanteren van een foutieve selectiemethode of het niet correct opvolgen van een theoretisch juiste selectiemethode kan leiden tot een steekproef waarin elementen zijn opgenomen die ten onrechte een grotere of kleinere kans hebben gehad dan andere elementen. Wanneer de samenstelling van het kader niet correct is (dubbeltellingen en omissies), wordt een dergelijk verschil in kans ook al verkregen. We zullen in deze deelparagraaf een onderscheid maken tussen het trekken van een steekproef uit een persoonsregister en het trekken van een steekproef op een adres (bijvoorbeeld uit huishoudens, afgiftepunten of uit personen). Het grote probleem van een steekproef die op een verkeerde manier is verkregen, is dat indien er een verband bestaat tussen de reden waarom de onderlinge kansen van de elementen verstoord zijn en het te onderzoeken verschijnsel, het nauwelijks mogelijk is te corrigeren. In het algemeen zal men proberen de gerealiseerde steekproef te verkleinen of uit te breiden, opdat bijvoorbeeld de verdeling van de steekproef naar leeftijd en geslacht overeenkomt met de zelfde verdeling in de populatie. Dit geeft echter niet de garantie dat de steekproef nu correct is. Wanneer bijvoorbeeld bij de steekproeftrekking van een verouderd adressenbestand gebruik wordt gemaakt, dan zal het duidelijk zijn dat hoe men ook aan de verdeling van de steekproef schaaft, men toch te weinig personen in nieuwbouwwoningen in de steekproef heeft. Een voorbeeld van een verkeerde selectieprocedure is als volgt. Stel dat de doelpopulatie bestaat uit personen van 18 jaar of ouder in de Nederlandse bevolking en er een steekproef wordt getrokken uit de persoonsregisters bij de gemeenten. Stel tevens dat, indien een persoon wordt getrokken die jonger is dan 18 jaar (een geval van overdekking), de eerstvolgende persoon in het register van 18 jaar of ouder wordt genomen. Dit kan leiden tot een te grote kans van bijvoorbeeld vaders. Aangezien de ordening in het persoonsregister alfabetisch is en kinderen overwegend de naam van de vader krijgen, wordt bij het gebruik van deze methode de kans om een vader in de steekproef te trekken vergroot. Indien er een verband bestaat tussen de gezinsverhouding of burgerlijke staat en de doelvariabelen van het onderzoek, wordt het onderzoek op deze wijze vertekend.

3.6.2. *Het trekken van een personensteekproef bij de gemeenten*

Indien men de keuze heeft gemaakt tussen een personensteekproef en een adressensteekproef (zie ook 3.4) en de keuze is gevallen op een personensteekproef, dan kan men de gemeenten vragen een steekproef te trekken. Dit is de werkwijze die door het CBS wordt gevolgd. Voorbeelden van steekproefonderzoek die op een personensteekproef zijn gebaseerd, zijn de Woningbehoeftenonderzoeken, de Vakantie-onderzoeken en de Leefsituatie-onderzoeken. Een probleem bij het geven van richtlijnen voor de gemeente-ambtenaren hoe de steekproef te trekken, wordt gevormd door de aanwezigheid van hulpbestanden waar de steekproef ook uit getrokken kan worden. Naast het persoonsregister is er vaak een hulpbestand waarmee allerlei werkzaamheden kunnen worden uitgevoerd. Dit hulpbestand kan volledig geautomatiseerd zijn (magneetbanden bij (regionale) computercentra), gedeeltelijk geautomatiseerd zijn (ponskaartenbestand bij de gemeente, dat periodiek naar een automatiseringscentrum wordt verzonden om te worden bijgewerkt) of mechanisch zijn (adresplaatjes waarvan een afdruk kan worden gemaakt). Indien de steekproeftrekking systematisch geschiedt en er een zinvolle ordening in het bestand aanwezig is, bijvoorbeeld naar leeftijd, dan heeft men de zekerheid dat de verschillende leeftijdscategorieën goed vertegenwoordigd zijn in de steekproef. Soortgelijke effecten treden op indien het hulpbestand gerangschikt is naar kieskring of wijk.

Bij het trekken van de steekproef uit het persoonsregister zelf, wat met name bij de kleinere gemeenten vaak gebeurt, kunnen er zich problemen voordoen door de eventuele aanwezigheid van niet-persoonskaarten (hulpkaarten). Veel gemeenten hebben behalve de persoonskaart van personen verwijs- of archiefkaarten aangelegd. Dit zijn hulpkaarten waarmee het persoonsregister wordt aangevuld en die door de gemeenten worden aangelegd voor eigen gemak. Verwijskaarten zijn 'gekopieerde' persoonskaarten die verwijzen naar de echte persoonskaart elders in de bak. Door een afwijkende kleur kunnen ze niet met de eigenlijke persoonskaart worden verward. Zo kan voor een gehuwde vrouw een verwijskaart worden gemaakt op de achternaam van haar echtgenoot. Deze vrouw komt dan twee keer in het register voor, één keer via haar persoonskaart en één keer via een verwijskaart in de buurt van de persoonskaart van haar man. Archiefkaarten worden aangelegd als personen zijn overleden of verhuisd naar een andere gemeente. Ook deze kaarten zijn louter voor het gemak van de gemeente. Ze kunnen staan op de plaats waar eerst de persoonskaart stond, dus in het persoonsregister, maar ook komt het voor dat gemeenten de archiefkaarten apart naast het persoonsregister bewaren. In het algemeen zal men steekproeven op systematische wijze trekken. Men bepaalt de stapgrootte S als inverse van de trekkingskans en kiest een aselect begin. In principe zouden de persoonskaarten moeten worden geteld, zodat zich steeds precies $S-1$ persoonskaarten tussen twee getrokken kaarten bevinden. Vaak zullen gemeenten niet tellen maar met behulp van een liniaal (ongeveer) S kaarten meten. Op zich is hier niets tegen, maar wel dient men zich er rekenschap van te geven dat oudere persoonskaarten, behorende bij oudere mensen, nu een grotere trekkingskans hebben. De oorzaak hiervan is dat kaarten naarmate ze ouder worden aan de bovenkant gaan slijten en wollig, rafelig worden. Hierdoor neemt het oppervlak aan de bovenkant van de kaart toe.

Een mogelijkheid om na te gaan of de steekproef goed is getrokken is het hanteren van een chi-kwadraattoets op de populatie-aantallen uitgesplitst naar bijvoorbeeld leeftijd en geslacht en de hiermee corresponderende steekproefaantallen. Uit de waarde van de chi-kwadraatgrootheid bij een zeker aantal vrijheidsgraden kan men de overschrijdingskans afleiden. Zo is er bij het Woningbehoeftenonderzoek 1977/'78 voor alle getrokken gemeenten nagegaan of de overschrijdingskans van de berekende chi-kwadraat acceptabel was. Overigens biedt een niet-significante chi-kwadraat geen garantie voor een goede steekproef.

3.6.3. *Het trekken van een steekproef op een adres*

Het trekken van een steekproef op een adres kan in twee fasen gebeuren. In de eerste plaats is het mogelijk dat er op een adres een steekproef uit huishoudens moet plaatsvinden en in de tweede plaats moeten, indien slechts een beperkt aantal personen per huishouden worden ondervraagd, de te ondervragen personen worden aangewezen. De fouten die hierbij kunnen optreden zijn het verkeerd selecteren van de te ondervragen persoon/personen door de enquêteur/trice en het niet volledig of correct inventariseren van de personen die op het adres wonen. Dit laatste probleem leidt tot onderdekking, hetgeen uitvoerig aan de orde is geweest in deelparagraaf 3.5. Het moet mogelijk zijn om de eerste foutenbron, een verkeerde selectie, onder controle te houden door voldoende aandacht te besteden aan de vele verschillende mogelijkheden die zich in het veld kunnen voordoen. Het is in de praktijk moeilijk na te gaan of de steekproef op het adres correct is getrokken. Om een significante afwijking te kunnen aantonen moet men in ieder geval de beschikking hebben over een groot aantal vraaggesprekken per enquêteur/trice.

3.7. *Het veldwerk*

3.7.1. *Inleiding*

Een groot aantal foutenbronnen kunnen bij het veldwerk optreden. We zullen hieronder de belangrijkste noemen. De hierbij gehanteerde volgorde geeft niet aan dat een verschijnsel een grotere of kleinere invloed heeft dan een later genoemd verschijnsel. Men kan de volgende foutenbronnen onderscheiden:

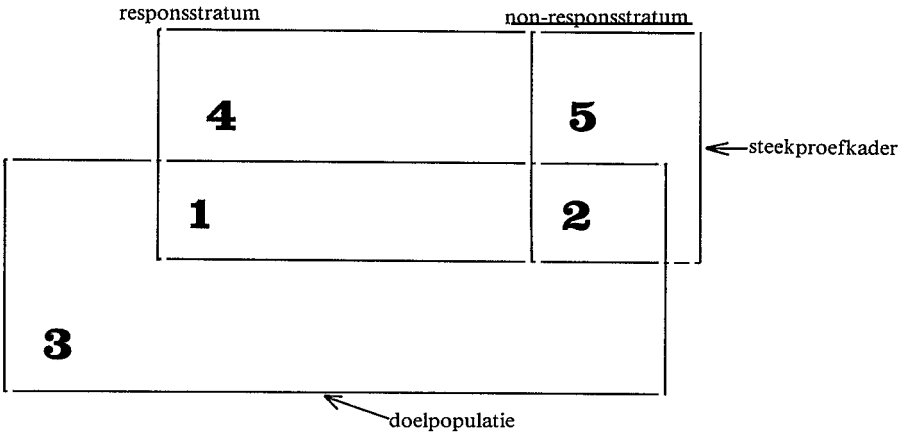
1. non-respons
2. meetfouten
3. intervieweffecten
4. herinneringseffecten
5. gewenningseffecten
6. 'organisatie-van-het-veldwerk'-effecten.

De hierboven gegeven opsomming is niet volledig één-éénduidig. Men kan voorbeelden bedenken die men onder twee of misschien wel meer categorieën kan indelen. Zo geldt bijvoorbeeld dat intervieweffecten aanleiding kunnen geven tot meetfouten, maar niet iedere meetfout hoeft het gevolg te zijn van een intervieweffect. We zullen in de deelparagrafen 3.7.2 - 3.7.7 ingaan op deze verschillende foutenbronnen.

3.7.2. *Non-respons*

Behalve door het CBS wordt of is hiernaar o.a. onderzoek gedaan door 't Hart (1974), Van Vliet (1981) en Nederhof (1981). Bij het CBS valt het non-respons-onderzoek in twee delen uiteen: onderzoek naar het tegengaan van non-respons in het veld en onderzoek naar correctiemethoden voor de invloed van non-respons. Widdershoven en Van den Berg (1980) rapporteren over onderzoek dat verricht is naar het beste tijdstip waarop een persoon kan worden benaderd voor een vraaggesprek. Ook wordt onderzoek verricht naar de enquêtedruk in Nederland en enquêtemoeheid. Bij het onderzoek naar de invloed van de aanschrijfbrief op het responspercentage is geëxperimenteerd met verschillende soorten brieven: korte en lange brieven, zakelijke en minder zakelijke brieven, brieven met en zonder voorbeelden uit de vragenlijst, brieven met en zonder telefoonnummer van de enquêtrice, enz. Resultaten van dit onderzoek worden gegeven door Van den Berg en Wetzels (1981). Onderzoek naar de invloed van non-respons op de uitkomsten van het steekproefonderzoek wordt op de hoofdafdeling Statistische Methoden van het CBS uitgevoerd. Resultaten van dit onderzoek zijn o.a. te vinden in Bethlehem en Kersten (1982b).

Een van de problemen van het non-responsonderzoek is de afbakening van het begrip non-respons. Zo is er een verschil tussen dekkingsfouten en non-respons. Dit onderscheid is van belang, aangezien bij de ophoging voor non-respons en dekkingsfouten op een verschillende manier kan worden gecorrigeerd. Het schema in figuur 2 kan worden uitgebreid door de populatie die via het steekproefkader kan worden bereikt, te verdelen in twee strata: het responsstratum en het non-responsstratum. Elementen die tot het responsstratum behoren zullen, wanneer ze in de steekproef worden getrokken, responderen, terwijl elementen uit het non-responsstratum dit niet zullen doen. Figuur 3 geeft dit schematisch weer.



Figuur 3. Schematische voorstelling van doelpopulatie, steekproefkader en respondenten

Er kunnen zich verschillende mogelijkheden voordoen wanneer we de elementen van de doelpopulatie vergelijken met de elementen van het steekproef-

kader en het respons- en non-responsstratum:

1. Hier bevinden zich elementen die tot de doelpopulatie behoren, die opgenomen zijn in het steekproefkader en die responderen. We kunnen dit *echte respons* noemen.
2. Dit zijn elementen die tot de doelpopulatie behoren, in het steekproefkader zijn opgenomen, maar niet responderen. Dit is de echte *non-respons*. Voorbeelden hiervan zijn weigering en niet-thuis.
3. Dit is *onderdekking*.

4 en 5 geven de *overdekking* aan. Met 4 komen elementen overeen die zouden responderen en met 5 elementen die niet zouden responderen. Bij contact met de elementen uit 4 zal blijken dat ze niet tot de doelpopulatie behoren en dan kunnen ze alsnog uit de steekproef worden verwijderd. De elementen uit 5 kunnen een probleem opleveren omdat wegens gebrek aan informatie het niet altijd mogelijk zal zijn om te onderkennen dat ze niet in de steekproef thuishoren, terwijl via een of andere correctiemethode aan deze elementen wèl een waarde kan worden toegekend.

3.7.3. *Meetfouten*

Deze fouten, die onder waarnemingsfouten vallen, worden veroorzaakt door het al dan niet opzettelijk geven van niet-correcte antwoorden. Hierbij speelt de formulering van de vragen, de volgorde e.d., in het algemeen aangeduid met de kwaliteit van het meetinstrument, een belangrijke rol. Men kan de kwaliteit van het meetinstrument bepalen met behulp van een test-hertestmethode voor dezelfde populatie. Onderzoek hiernaar wordt o.a. verricht door Van de Pol (1982).

3.7.4. *Interviewefffecten*

De interactie tussen de ondervraagde en enquêteur/trice speelt een belangrijke rol voor de kwaliteit van de gegevens. Afgezien van effecten die ontstaan door het uiterlijk van de enquêteur/trice, de sociale wenselijkheid van antwoorden, het niet-erkennen van het feit dat men de vraag niet begrijpt of er geen mening over heeft, kan de aanwezigheid van andere personen bij het vraaggesprek een rol van betekenis spelen. Onderzoek naar interviewefffecten wordt verricht door o.a. Van der Eijk en Niemöller (1981), Van der Zouwen (1981), 't Hart (1978) en Dijkstra en Elsinga (1979).

3.7.5. *Herinneringseffecten*

De mens onthoudt sommige gebeurtenissen beter dan andere gebeurtenissen. Voor steekproefonderzoek naar het optreden van bepaalde verschijnselen of het doen van bepaalde handelingen (bijvoorbeeld aankoopgedrag, bioscoopbezoek, raadplegen arts of het om deelname worden gevraagd aan een enquête) speelt dit effect van 'telescoping' een belangrijke rol. Vaak heeft een dergelijk onderzoek tot doel de frequentie te schatten waarmee binnen een zeker tijdsinterval een bepaald verschijnsel zich voordoet. Het herinneringseffect zal dan tot een te hoge schatting van de frequentie leiden indien personen een opgetreden gebeurtenis of handeling dichterbij in de tijd schatten dan deze in werkelijkheid plaatsvond. Voor bepaalde steekproefonderzoeken is het mogelijk om de antwoorden van respondenten te vergelijken met de werkelijke

gebeurtenissen. Demografisch en medisch steekproefonderzoek zijn hier voorbeelden van. Dit vergelijken van door respondenten gegeven antwoorden met doorgaans administratieve gegevens maakt het mogelijk om voor specifiek steekproefonderzoek het herinneringseffect te modelleren. Bailar en Biemer (1981) en Som (1973) hanteren een exponentieel model. Een dergelijk model blijkt in de praktijk van demografisch en medisch steekproefonderzoek goed te werken.

3.7.6. *Gewenningseffecten*

Onderzoeksgewenningseffecten (Engels: 'rotation group bias' of 'time in sample bias') treden op bij het gebruik van panels bij steekproefonderzoek. Na verloop van tijd kan men een zekere routine ontwikkelen bij het beantwoorden van de vragen. Daarnaast is het mogelijk dat iemand een foutief antwoord geeft om de continuïteit van de antwoorden te behouden dan wel juist te onderbreken. In Bailar en Biemer (1981) wordt onderzoek naar deze effecten beschreven.

3.7.7. *'Organisatie-van-het-veldwerk'-effecten*

De algemene organisatie van het veldwerk kan ook van invloed zijn op de kwaliteit van de uitkomsten. Men kan hierbij niet alleen denken aan het behalen van een zeker percentage respons, maar ook aan de invulling van de vragenlijsten (partiële non-respons en de consistentie van de antwoorden). Uit de CBS-praktijk blijkt dat het geven van een mondelinge instructie aan de enquêtrices kan leiden tot een aanmerkelijk hogere respons. Eenzelfde effect wordt verkregen door meer richtlijnen te geven over de tijdstippen waarop de kans op thuistreffen groter is. Dit kan overigens leiden tot een selectere respons. Widdershoven en Van den Berg (1981) doen verslag hierover. De veldwerkorganisatie heeft niet alleen invloed op het totale percentage non-respons, maar kan ook invloed hebben op de selectheid ervan. Dekkingsproblemen kunnen behalve bij het trekken van de steekproef ook optreden bij het veldwerk. Door een goede instructie van de enquêteurs/trices en het correct opvolgen van deze instructie kunnen deze problemen worden opgelost en geven ze geen aanleiding tot vertekening.

3.8. *De ophoging*

Onder de ophoging (het ophogen) verstaat men het herwegen van de uitkomsten van de enquête om uitspraken te kunnen doen over de gehele doelpopulatie. De term ophogen wordt meestal gehanteerd bij het schatten van totalen (bijvoorbeeld het aantal woningzoekenden in Nederland, het totale aantal uitstapjes van de Nederlandse bevolking, enz.). De ophoging biedt de mogelijkheid eventuele fouten te corrigeren. Hierbij kan gebruik worden gemaakt van hulpinformatie van de populatie (bijvoorbeeld de verdeling van de Nederlandse bevolking naar leeftijd en geslacht) of van hulpinformatie betreffende de respondenten en non-respondenten. De meest bekende techniek die bij het ophogen wordt gehanteerd is stratificatie achteraf. Een overzicht van methoden die met name tot doel hebben de vertekening ten gevolge van non-respons terug te dringen wordt gegeven in Bethlehem en Kersten (1982c). Het is echter

ook mogelijk dat bij het ophogen juist een extra vertekening wordt geïntroduceerd. Dit zal met name gebeuren indien de veronderstellingen die aan de ophoging ten grondslag liggen niet geldig zijn. Vandaar dat het onderzoek naar mogelijke fouten bij steekproefonderzoek zo nodig is.

4. SLOT

In de voorgaande paragrafen is een uiteenzetting gegeven van de verschillende foutenbronnen bij steekproefonderzoek. Gezien de relatie tussen de totale fout, de steekproeffout en de niet-steekproeffout is onderzoek naar de niet-steekproeffout ten zeerste van belang. In verschillende landen, vooral in de Verenigde Staten maar ook in Nederland, wordt onderzoek hiernaar verricht en het is te hopen dat publicaties over steekproefonderzoek binnen afzienbare tijd niet alleen informatie bevatten over de steekproeffout maar ook over de grootte en richting van de niet-steekproeffouten.

BIBLIOGRAFIE

Aigner, D.J., 1979. *Sample design for electricity pricing experiments,* Journal of Econometrics, Vol 11, pp. 195-205.

Bailar, B.A. en P.P. Biemer, 1981. *Some methods for evaluating nonsampling error in household censuses and surveys,* Report by the Bureau of the Census of the United States, presented at the Meeting on problems relating to household surveys of the U.N. Conference of European Statisticians, Geneva, 1-4 June.

Berg, J. van den, 1979. *Non-respons van het consumenten conjunctuuronderzoek,* Intern CBS-rapport (Centraal Bureau voor de Statistiek, Heerlen).

Berg, J. van den en W.A.C.H. Wetzels, 1981. *Aanschrijfbrieven en non-respons,* Intern CBS-rapport (Centraal Bureau voor de Statistiek, Heerlen).

Bethlehem, J.G. en H.M.P. Kersten, 1982a. *The nonresponse problem,* Survey Methodology, Vol. 7, pp. 130-156.

Bethlehem, J.G. en H.M.P. Kersten, 1982b. *De non-responsanalyse van de woningbehoeftenonderzoeken,* Statistisch Magazine, Vol 1, no. 3, pp. 80-96.

Bethlehem, J.G. en H.M.P. Kersten, 1982c. Non-respons bij enquêtes, of 'Wat niet weet, wat niet deert?', Mens en Maatschappij, Vol. 57., pp. 145-170.

Deighton, R.E., J.R. Poland, J.R. Stubbs en R.D. Tortora, 1978. *Glossary of nonsampling error terms: an illustration of a semantic problem in statistics,* Statistical Policy Working Paper 4 (U.S. Department of Commerce).

Dijkstra, W. en E. Elsinga, 1979. *Some determinants of interview bias,* Methoden en Data Nieuwsbrief, Vol 4, no. 3, pp. 91-118.

Eijk, C. van der en B. Niemöller, 1981. *Interview effects. The emperor's new clothes?,* Handelingen van het congres van de Dutch Sociometric Society, pp. 20-36.

Hart, H. 't, 1974. *Selectie en zelfselectie van informanten in enquêtes,* proefschrift Universiteit van Amsterdam.

Hart, H. 't, 1978. *Andere aanwezigen bij het vraaggesprek,* In: Röntgenfoto van Enquête-onderzoek, pp. 76-99. Werkgroep kwartaalonderzoek, Universiteit van Amsterdam.

Kish, L., 1967. *Survey sampling,* John Wiley & Sons, Inc., New York.

Kersten, H.M.P. en J.G. Bethlehem, 1981. *Ontwikkelingen op het gebied van non-responsanalyse bij het CBS,* Handelingen van het congres van de Dutch Sociometric Society, pp. 222-256.

Murthy, M.N., 1979. *A framework for studying incomplete data with a reference to the experience in some countries of Asia and the Pacific.* In: Symposium on incomplete data: preliminary proceedings, U.S. Department of Health, Education and Welfare.

Nederhof, A.J., 1981. *Some sources of artifact in social research: nonresponse, volunteering and research experience of subjects,* Proefschrift Rijksuniversiteit Leiden.

Pol, F.J.R. van de, 1982. *Betrouwbaarheidsbepaling met herhaalde metingen,* Mens en Maatschappij, Vol. 57, pp. 70-95.
Ree, S.J.M. de en N.J. Nieuwenbroek, 1981. *Precision of various estimators for simple random and single stage cluster samples,* Paper presented at the Meeting on problems relating to household surveys of the U.N. Conference of European Statisticians, Geneva, 1-4 June.
Som, R.K., 1973. *Recall lapse in demographic enquiries,* Asia Publishing House, New York.
Vliet, J. van, 1981. *Nonresponse en onderzoekkwaliteit,* Paper ten behoeve van de dag over 'Kwaliteit van Mondelinge Enquêtes' georganiseerd door de Nederlandse Vereniging van Marktonderzoekers.
Widdershoven, M. en J. van den Berg, 1980. *Non-respons bij twee 'persoons- en gezinsenquêtes',* In: CBS-Select 1, pp. 357-366, Staatsuitgeverij, s'-Gravenhage.
Zouwen, J. van der, 1981. *Masia, een simulatiemodel van het beantwoorden van vragen in survey-interviews,* Handelingen van het congres van de Dutch Sociometric Society, pp. 54-63.

5. Mediagebruik 1975-1980 aan de hand van tijdsbestedingsonderzoek

W.P. KNULST

1. INLEIDING

Dit artikel is gebaseerd op gegevens, ontleend aan twee tijdsbestedingsonderzoeken verricht in 1975 en 1980 en met grotendeels overeenkomstige vraagstellingen. Om na te gaan op welk tijdstip radio en televisie worden gebruikt, heeft de Nederlandse Omroep Stichting in 1970 een onderzoek laten houden naar de gehele tijdsbesteding gedurende één week.[1] Het onderzoek werd verricht door Intomart B.V. te Hilversum. Toen in 1975 dit laatstgenoemde onderzoek door Intomart zou worden herhaald, traden behalve de N.O.S., het Sociaal en Cultureel Planbureau en enige andere instellingen als opdrachtgever op.

De tijdsbesteding zou worden onderzocht aan de hand van een uitgebreide lijst van activiteiten, waarbij in grote lijnen de rubricering zou worden aangehouden welke is gebruikt bij het cross-national tijdsbestedingsonderzoek uit 1965-'66, dat onder auspiciën stond van het European Coordination Centre for Research and Documentation in Social Sciences te Wenen[2]. Tevens zou de tijdsbesteding niet per halfuur, zoals in 1970, maar per kwartier worden geregistreerd. In de periode van 5 tot en met 19 oktober 1975 heeft het tijdsbestedingsonderzoek plaatsgevonden onder een representatieve groep van ca. 1300 Nederlanders van 12 jaar en ouder. De tijdsbesteding werd gedurende één week geregistreerd. De respondent werd gevraagd zijn bezigheden van kwartier tot kwartier te scoren met behulp van een gestructureerde lijst van ca. 200 activiteiten, verdeeld over 10 hoofdgroepen. Voorafgaande aan de registratie van hun tijdsbesteding zijn de respondenten over verschillende zaken geënquêteerd. Ter afsluiting van de week waarin het dagboek is bijgehouden, is er een kleinere enquête gehouden.[3] Een zo volledig mogelijke replicatie van het project heeft in 1980 plaatsgevonden. Ook in 1980 is het onderzoek in de periode van 5 tot en met 19 oktober gehouden, maar nu onder ca. 2700 Nederlanders van twaalf jaar en ouder.

In de volgende paragrafen wordt mediagebruik als bestanddeel van het tijdsbudget geanalyseerd. Hierdoor kon alleen gelet worden op mediagebruik als *primaire* activiteit. De gevolgen hiervan zijn dat de uitspraken over met name radiogebruik beperkte betekenis hebben, omdat vastgesteld is dat 94% van het radiogebruik in 1980 als secundaire activiteit voorkomt. Bij televisiegebruik is 26% gescoord als secundaire activiteit.

2. *Enige algemene verschuivingen tussen 1975 en 1980*

Concentreert men zich op verschuivingen in tijdsbesteding 1975-1980 (tabel 1, 2, appendix) dan zijn er de volgende, meest opvallende feiten:

- De gemiddelde arbeidstijd van werkende mannen uit 1980 is ca. 80 minuten minder dan die van werkende mannen uit 1975. Deze gewonnen tijd kwam niet terecht bij de vrijetijdsbesteding – de tijdsbesteding voor media en overige vrijetijdsbesteding verminderde met ruim één uur – maar verschoof naar werk voor de huishouding en de zorg voor het gezin.
- Bij werkende vrouwen zijn er geen opmerkelijke verschuivingen in de tijdsbesteding aan arbeid of huishouding. Opmerkelijk is een toename van het mediagebruik ten koste van andere vrijetijdsbezigheden. Het verschil in tijdsbesteding aan media tussen werkende mannen en werkende vrouwen blijft aanzienlijk, maar was in 1980 geringer dan in 1975.
- De scholieren en studenten uit het onderzoek 1980 besteedden bijna 2 uur minder aan hun studie dan de scholieren en studenten uit het onderzoek 1975. De vrije tijd was bij de ondervraagde studenten/scholieren uit 1980 bijna evenveel meer als de studietijd minder was.
- Gepensioneerde mannen uit het onderzoek 1980 besteedden nog weer meer tijd aan media dan die in 1975. Met 30 uren per week vullen zij bijna de helft van hun vrije tijd met media.

2.1. *Mediagebruik (als primaire activiteit) 1975-1980 nader bezien*

De tijdsbesteding aan mediagebruik onder de gehele bevolking vertoont tussen 1975 en 1980 geen opvallende verandering. Bovendien bleef de tijdsverdeling over de verschillende electronische en gedrukte media in grote lijnen gelijk (tabel 3 en tabel 4). Verschuivingen traden wel op in het volume van mediagebruik binnen categorieën uit de bevolking (tabel 4):

- Mannen vergrootten hun overwicht in tijdsbesteding aan electronische media en kranten op vrouwen. De voorsprong in tijdsgebruik voor tijdschriften en boeken door vrouwen, werd minder.
- Met uitzondering van muziekplaten/banden besteden ouderen meer tijd aan media dan jongeren. Meer convergentie trad echter op in tijdsbesteding aan electronische media door de verschillende leeftijdsgroepen. De verschillen tussen de leeftijdsgroepen in tijdsbesteding aan gedrukte media werden groter. De afgenomen belangstelling voor dagbladen en tijdschriften onder jongeren valt hierbij wel bijzonder op.
- In het algemeen besteden personen met een laag onderwijsniveau meer tijd aan electronische media dan hoger opgeleiden. Deze laatsten steken meer tijd in gedrukte media dan laag opgeleiden. Convergentie in tijdsbesteding door de verschillende opleidingsgroeperingen trad op bij televisie, muziekplaten en -banden, kranten en tijdschriften.

3. TIJDSBUDGETCONSTRAINTS EN MEDIAGEBRUIK

3.1. *Arbeidstijd en mediagebruik*

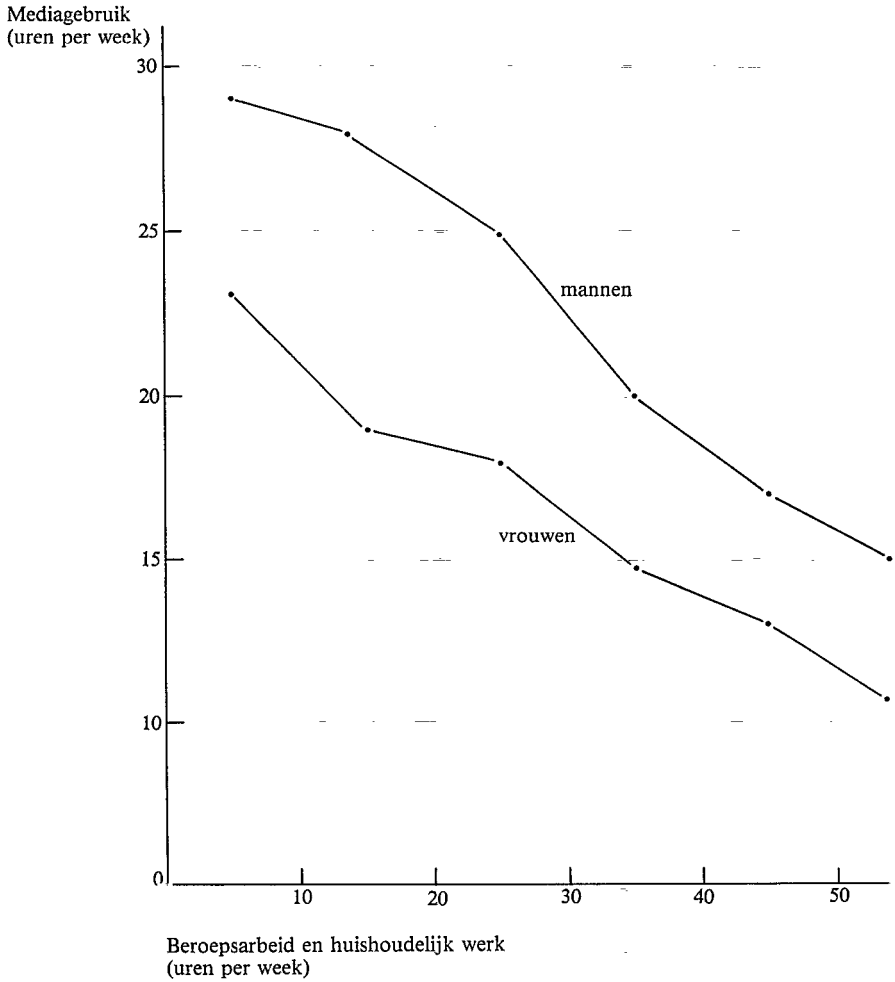
In de vorige paragraaf werd aangetoond dat personen met een werkkring buitenshuis of huisvrouwen die de zorg voor de huishouding dragen, minder uren investeren in media dan personen die geen vaste dagtaak hebben, zoals gepensioneerden. Het mediagebruik varieert van 32% van de vrije tijd bij werkende vrouwen – die veelal naast een beroep de zorg voor de huishouding hebben – tot 46% van de vrije tijd bij gepensioneerde mannen, die behalve assistentie in de huishouding geen vaste verplichtingen hebben. Thans wordt op systematische wijze ingegaan op de relatie arbeid-mediagebruik.

In tabel 5 is de tijdsbesteding voor verschillende soorten media (televisie, radio, muziekplaten en -tapes, boeken, kranten en tijdschriften) in verband gebracht met het gemiddelde aantal uren arbeid per week.

Hiertoe zijn alle respondenten – met uitzondering van schoolgaanden of studerenden – verdeeld naar het aantal uren arbeid (incl. bezigheden in verband met beroepsarbeid) dat zij scoorden per week.

- Het totale mediagebruik loopt bij mannen af van 28,4 uur bij 0 uur arbeid tot 13,8 uur bij 51 en meer uur arbeid per week. Bij vrouwen is er een minder groot verloop: van 15,9 uur (bij 0 uur arbeid) tot 10,2 uur (bij 51 en meer uur arbeid per week).
- Bij mannen bestaat tussen 0 en 20 uur arbeid een praktisch elastische aanpassing van mediagebruik. Tussen 21 en 50 uur arbeid reageert mediagebruik nauwelijks elastisch op het aantal gewerkte uren, maar daarboven wordt de elasticiteit weer groter. Bij vrouwen bestaat een weinig elastische aanpassing van mediagebruik aan het aantal uren arbeid per week.

Het verschil tussen mannen en vrouwen moet worden toegeschreven aan het feit dat arbeidstijd een betere predictor voor mannen vormt voor het vrijetijdsvolume dan bij vrouwen. Tussen arbeid en vrije tijd bestaat bij vrouwen nog een aanzienlijke verplichting op het gebied van de huishouding, wat bij mannen niet in die mate het geval is:
Vrouwen die 31 tot 40 uur per week besteden aan beroepsarbeid besteden aan arbeid en huishouding samen gemiddeld 51 uur; mannen die 31 tot 40 uur per week werken, besteden aan beroepsarbeid en huishouding samen gemiddeld 42 uur. Het verschil in mediagebruik verdwijnt echter niet indien men mediagebruik in verband brengt – niet met arbeid sec – maar met het totaal van arbeid èn huishoudelijk werk (grafiek 1). In paragraaf 3.5 wordt verder ingegaan op een andere factor die van invloed is op het verschil in mediagebruik tussen mannen en vrouwen. De conclusie luidt dat onder de huidige ongelijke verdeling van verplichte bezigheden over mannen en vrouwen, de omvang van de arbeidstijd alleen bij mannen een geschikte predictor is voor het aantal uren mediagebruik. Op grond van de gevonden resultaten kan men stellen dat wijzigingen in arbeidstijd in de zone tussen 40 en 20 uur, weinig gevolgen hebben voor mediagebruik. Een eventuele officiële arbeidstijdverkorting van bijv. 40 naar 35 uur per week zal nauwelijks gevolgen hebben voor mediagebruik in zijn totaliteit en evenmin voor het gebruik van afzonderlijke media. Verschui-



Grafiek 1. Tijdsbesteding aan mediagebruik naar gewerkte uren (arbeid + huishouding) voor mannen en vrouwen (12 jaar en ouder m.u.v. studerenden, oktober 1980).

ving echter in het totale pakket voor verplichte bezigheden (herverdeling van arbeid en huishoudelijk werk over mannen en vrouwen) zou waarschijnlijk wel waarneembare gevolgen hebben voor mediagebruik.

Bovendien hebben drastische wijzigingen in arbeidstijd, zoals bij werkloosheid of pensionering, aanzienlijke gevolgen voor mediagebruik, zoals uit paragraaf 3.3 moge blijken.

3.2. *Omvang vrije tijd en de hoeveelheid mediagebruik*

De omvang van de vrije tijd (= tijd die resteert na aftrek van tijdsbesteding aan arbeid, huishouding, onderwijs, slapen, eten en persoonlijke verzorging) laat zich moeilijker vaststellen dan de gemiddelde arbeidstijd. De gemiddelde arbeidstijd van verschillende groepen werknemers wordt regelmatig gere-

gistreerd; voor berekening van de vrije tijd is tijdsbudgetonderzoek nodig. Nochtans geeft kennis van de ontwikkeling van het vrijetijdsvolume de meest betrouwbare basis voor uitspraken over mediagebruik. Tabel 6 leert dat het totale mediagebruik gemiddeld 34 tot 40% van de vrije tijd uitmaakt en praktisch evenredig groeit met een volumetoename van de vrije tijd. De daling in de mediaproportie bij een toename van de vrije tijd is te gering om er conclusies aan te verbinden.

3.3. *Mediagebruik bij werkende en niet-werkenden*

Hierboven werd vastgesteld dat mediagebruik (bij mannen) zeer elastisch reageert in de zône van 0 tot 20 uur arbeid per week. Elk uur dat in die zone minder (of meer) wordt gewerkt gaat gepaard met een verschuiving van bijna één uur in mediagebruik.

Deze schets van de situatie is natuurlijk vrij theoretisch. Personen die in een situatie komen te verkeren waarin ze niet langer werken, hebben in de meeste gevallen geen werkkring van 1 tot 20 uur maar een werkkring van ca. 40 uur per week verlaten.

Voor een beoordeling van de gevolgen van pensionering of werkloosheid voor het mediagebruik kan men daarom beter de situatie van werkende mannen die meer dan 20 uur per week werken, vergelijken met die van niet-werkenden.

De vergelijking toont aan dat het mediagebruik bij niet-werkende mannen (gepensioneerden èn werklozen) disproportioneel sterker toeneemt dan hun tijdsbesteding aan overige bezigheden (met uitzondering van huishoudelijk werk). Vergelijkt men het mediagebruik van werkende vrouwen met dat van huisvrouwen (tabel 2) dan leert zulks, dat verschuivingen van werk buitenshuis naar werk in de huishouding nauwelijks gevolgen heeft voor mediaconsumptie, althans in 1980.

Men kan concluderen dat de toename van het aantal niet-werkende (gepensioneerden èn werklozen) veel drastischer gevolgen heeft voor individueel mediagebruik dan effecten van arbeidsdeelname door vrouwen of gevolgen van arbeidstijdsverkorting. Aan het effect van de groei van het aantal werklozen en gepensioneerden moet nog dat van de autonome stijging in mediagebruik onder niet-actieven tussen 1975 en 1980 worden toegevoegd (vgl. tabel 2).

3.4. *Mediagebruik en verscheidenheid in vrijetijdsbesteding*

Mediagebruik blijkt beslag te leggen op een tamelijk constant deel van de vrije tijd, hetgeen erop wijst dat de omvang van de vrije tijd binnen het individuele tijdsbudget een belangrijke factor vormt voor de verdeling van tijd over media en andere vrijetijdsactiviteiten. De proportie mediagebruik is echter een gemiddelde en het feit dat werkende vrouwen 32% van hun vrije tijd aan media besteden tegenover gepensioneerde mannen 46%, vraagt een aparte verklaring.

In deze en de volgende paragraaf zijn daarvoor twee andere budgetcondities onderzocht: de mate van verscheidenheid van het vrijetijdsbudget en de mate

van versnippering van het vrijetijdsbudget.

In verschillende studies is opgevallen dat personen die zeer intensief gebruik maken van televisie, gemiddeld minder andere liefhebberijen hebben dan personen die weinig televisie kijken[4].

Analyses van tijdsbesteding van niet-werkenden in vergelijking met werkende mannen leren dat een zeer intensief mediagebruik van niet-werkende mannen gepaard gaat met een weinig gedifferentieerd vrijetijdsleven op andere terreinen[5]. Vrouwen met een betaalde baan die in 1975 beduidend lager scoorden op mediagebruik dan huisvrouwen, vertoonden een afwisselender vrijetijdspatroon dan huisvrouwen[6]. De veronderstelling lijkt gewettigd dat een intensief mediagebruik niet noodzakelijk samenhangt met een intensieve interesse in media, maar mogelijk ook met de afwezigheid van alternatieve mogelijkheden voor vrijetijdsbesteding. Deze hypothese is onderzocht binnen de termen van de budgetstudie. (Gegevens over de individuele beleving zijn niet beschikbaar.) In tabel 7 is de relatie tussen de verschillende vormen van mediagebruik en de verscheidenheid van de overige vrijetijdsbesteding geanalyseerd.

Met uitzondering van het gebruik van muziekplaten/-banden, tijdschriften en kranten, zijn alle overige relaties tussen mediagebruik en de verscheidenheidscore significant (op 5% niveau),
en wel als volgt:

- personen met een weinig afwisselend vrijetijdsleven maken intensiever gebruik van televisie ($r = -.23$) en intensiever gebruik van radio ($r = -.08$),
- personen echter met een afwisselend vrijetijdsleven scoren significant hoger op boekenlezen ($r = +.12$).

3.5. *Versnippering vrijetijdsbudget en mediagebruik*[7]

Vrouwen besteedden zowel in 1975 als in 1980 een geringer deel van hun vrijetijdsbudget aan media dan mannen (het verschil is in 1980 toegenomen: zie tabel 2; het verschil tussen vrouwen met werkkring en mannen met werkkring is zelfs nog groter). Een verklaring hiervoor is gezocht in de veronderstelling dat vrouwen onder de heersende rolverdeling over minder aaneengesloten blokken vrije tijd beschikken dan mannen, omdat hun vrijetijdsbezigheden worden afgewisseld met resp. verstoord door verplichtingen aan huishouding en kinderverzorging. Op grond van analyses van de budgetgegevens uit 1975 blijkt die veronderstelling juist te zijn:

- Werkende mannen in een gezin waarvan het jongste kind 0 tot en met 5 jaar oud is, beschikken per week over gemiddeld 26 blokken ononderbroken vrijetijdsbesteding; werkende vrouwen of huisvrouwen in een gezin met een jongste kind in de leeftijd van 0 tot en met 5 jaar beschikken gemiddeld over 54 blokken vrije tijd per week. Vrouwen in vergelijkbare gezinsomstandigheden als mannen, ongeacht of zij naast de huishouding een betaalde werkkring hebben, zien hun vrijetijdsbezigheden 2 maal vaker per week onderbroken door verplichtingen.
- De tegenstelling is ook – zij het minder schril – aanwezig tussen mannen en vrouwen die geen kleine kinderen hebben. Controleert men bovendien voor de mogelijke invloed van verschil in vrijetijdsvolume, opleiding en inkomen, dan blijft er een sterk significant verband tussen sexe en versnippe-

ring van de vrije tijd bestaan.

- Dit verschil is relevant, omdat er tevens een sterk verband werd geconstateerd tussen de mate van versnippering en het gebruik van televisie, radio, afspeelapparatuur èn kranten: hoe meer aaneengesloten de vrije tijd, hoe meer gebruik van de genoemde media.
- Naar aanleiding van deze bevindingen rees de vraag of mannen hoger scoren op mediagebruik dan vrouwen door een grotere mate van aaneengeslotenheid van hun vrije tijd of doordat mannen meer interesse hebben voor media.
- Bij analyse wees uit dat het verschil in gebruik van electronische media (onder controle van 9 mogelijke interveniërende verbanden) tussen mannen en vrouwen kan worden toegeschreven aan de verschillende structuren van hun vrije tijd. Voor krantenlezen is dit niet het geval. Hier blijft het verband met sexe significant (onder controle van interveniërende verbanden).

Op grond van deze analyse kan de relatie als volgt worden geschematiseerd:

geslacht ——→ verschil in aaneengeslotenheid vrije tijd ——→ verschil in tijdsgebruik electronische media

4. RELATIE TELEVISIEGEBRUIK – GEBRUIK ANDERE MEDIA

4.1. *Inleiding*

Er is vaak verondersteld dat de televisie andere vormen van vrijetijdsbesteding zou beconcurreren of zelfs verdringen. Op deze algemene manier gesteld is de veronderstelling niet juist.

De ontwikkeling in vrijetijdsbesteding heeft invloed ondergaan van meer factoren dan uitsluitend televisie, terwijl de televisie voor sommige vormen van ontspanning of ontplooiing nauwelijks relevant lijkt (toerisme, sport, hobby-beoefening e.d.). Wel is vastgesteld – en dit verschijnsel is in het vorige hoofdstuk ter sprake gekomen – dat personen die een meer dan gemiddeld aantal uren televisie kijken of radio beluisteren, een minder veelzijdig vrijetijdspatroon hebben dan personen die deze media minder gebruiken. Ook hier kan men niet zonder meer het mediagebruik als oorzaak aanwijzen, omdat er achterliggende factoren zijn die zowel de mate van televisiegebruik als de veelzijdigheid van de vrijetijdsbesteding beïnvloeden (bijv. leeftijd, ontwikkelingsniveau).

Omdat aan televisie, radio, kranten en tijdschriften overeenkomstige functies worden toegeschreven (informatie, amusement e.a.), is de verleiding groot om een direct verband te veronderstellen tussen het gebruik van verschillende soorten media.

Het tijdsbudgetonderzoek uit 1975 en 1980 maakt het mogelijk om de verdere ontwikkeling van de relatie tussen televisiegebruik en gebruik van andere media te registreren. Hierbij zijn twee aspecten interessant:

- de verhouding tussen de intensiteit van televisiegebruik en het gebruik van andere media op één tijdstip;

– de verhouding tussen intensiteit van televisiegebruik en gebruik van andere media in 1975 èn 1980.

4.2. *Verhouding televisiegebruik en gebruik andere media op één tijdstip*

In tabel 8 is de relatie tussen intensiteit van televisiegebruik en het gebruik van andere media geanalyseerd voor de week van oktober 1980. Hieruit blijkt dat een toenemende intensiteit van televisiegebruik samengaat met een toenemende intensiteit van radiogebruik (als primaire activiteit).

Naarmate er meer uren naar televisie is gekeken, is de radio vaker object van aandacht geweest en ligt bovendien de tijd voor radioluisteren hoger.

Bij het beluisteren van muziekplaten/-banden loopt de gemiddeld bestede tijd op bij een intensiever televisiegebruik, maar niet de participatie.
Bij de andere media bestaat ongeveer dezelfde relatie. Het gebruikspercentage van gedrukte media hangt niet samen met de intensiteit van televisiegebruik. De zware televisiekijkers scoren echter evenals bij radio en platenspelen, meer tijd bij het gebruik van gedrukte media dan matige of lichte kijkers. Dit verschil kan worden verklaard door het feit dat zware televisiekijkers gemiddeld over meer vrije uren beschikken dan matige of lichte kijkers. Schakelt men dit effect uit en daarbij ook de mogelijke invloed van opleidings- en sexeverschillen, dan resteert er een significant positief verband tussen televisiegebruik en radiogebruik, terwijl de score bij andere media niet significant varieert met televisiegebruik.

4.3. *Televisiegebruik – gebruik verschillende soorten tijdschriften*

Het gebruik van gedrukte media in het algemeen moge geen concurrentie ondervinden van de mate van televisiegebruik, maar hoe ligt de relatie tussen intensiteit televisiegebruik en het gebruik van specifieke soorten van gedrukte media? De resultaten van deze analyse geven aanleiding om het aanvankelijke beeld te nuanceren:

Het gebruik van *vrouwenbladen* (evenmin als het gebruik van jongerentijdschriften zoals strip- en muziekbladen) hangt niet significant samen met de intensiteit van televisiegebruik, maar voor andere bladen ligt dit anders;

Hobby-tijdschriften en *opinieweekbladen* (politieke, economische en culturele onderwerpen) worden significant minder gebruikt door zware televisiekijkers dan door matige of lichte kijkers. (Hierbij is evenals bij het gebruik van andere bladen de invloed van verschil in vrije uren, opleidingsniveau en sexe gecontroleerd.) Een verklaring hiervoor is dat zware televisiekijkers zoals in hoofdstuk 3 is aangetoond, een minder veelzijdig vrijetijds- c.q. hobbypatroon hebben dan matige en lichte kijkers.

Voor de overige gevonden relaties ligt de verklaring nog meer voor de hand: een hogere intensiteit van televisiegebruik gaat (significant) samen met een intensiever gebruik van *omroepbladen* en *populaire bladen*.

Indien men bedenkt dat populaire bladen voor een belangrijk deel zijn gevuld met gossip over televisie- en showsterren dan is dit positieve verband niet zo opvallend.

(In tabel 9 treft men een overzicht aan. In deze tabel zijn de invloeden van

verschil in vrije tijd, opleidingsniveau en sexe *niet* gecontroleerd.)

Men kan concluderen dat een intensief televisiegebruik alleen negatief verband houdt met het gebruik van media die een meer specifieke interesse verraden: hobby-tijdschriften en opinieweekbladen. Daarbij is het opmerkelijk dat het lezen van boeken – dat men evenmin onder algemene interesses kan voegen – niet afneemt met een intensiever televisiegebruik.

4.4. *Relatie televisiegebruik en gebruik andere media in 1975 en 1980*

De verhoudingen die hierboven zijn beschreven, waren in 1975 niet geheel anders. De positieve relatie televisiegebruik - radiogebruik was zelfs nog iets sterker dan in 1975. Daarnaast is de verhouding televisiegebruik - gebruik van platenspeler-taperecorder, tijdschriften en kranten gewijzigd bij de lichte en matige kijkers. Het resultaat van die verschuiving levert echter – zoals we hierboven zagen – geen significant verband op in 1980 voor kranten en tijdschriften (samengerekend). De verschuivingen waren bovendien van zodanige aard, dat niet de concurrentie van televisie t.o.v. andere media sterker werd, maar het cumulatieve patroon in tijdsbesteding aan media.

Bij de vermelde resultaten dient men te bedenken dat deze zijn vastgesteld op tijdstippen dat de Nederlandse televisiezenders op werkdagen geen uitzendingen overdag verzorgden. Over enige tijd zal deze situatie zich wijzigen.
Het is daarom aannemelijk dat de verhoudingen zich in een toekomstige aanbodsituatie, waarin de televisie ook overdag regelmatig als concurrent voor de krant, het tijdschrift of het boek gaat optreden, gaan wijzigen. Afgaande op gegevens uit landen waar televisie permanent overdag beschikbaar is, zal die concurrentie hoofdzakelijk optreden bij personen die tijdens werkdagen in de gelegenheid zijn televisie te kijken: gepensioneerden, werklozen of huisvrouwen[9].

NOTEN

1. Tijdsbesteding in Nederland, N.O.S.; S.T.E.R./Intomart z.j.
2. A. Szalai (e.d.) 1972. The Use of time, Den Haag/Parijs.
3. Sociaal en Cultureel Planbureau, W.P. Knulst, 1977. Een week tijd, Den Haag.
4. N.O.S. Afd. Kijk- en luisteronderzoek/A.M. Overste, 1975. De betekenis van televisie(kijken) – Hilversum.
5. Sociaal en Cultureel Planbureau, W.P. Knulst, 1977, pp. 34-43.
6. Sociaal en Cultureel Planbureau, W.P. Knulst, 1977, pp. 44-52.
7. Alle gegevens in deze paragraaf afkomstig uit: Sociaal en Cultureel Planbureau, W.P. Knulst, 1977, pp. 89-92.
8. Sociaal en Cultureel Rapport 1976, 1977, p. 153.
9. Sociaal en Cultureel Planbureau, L.P.H. Schoonderwoerd en W.P. Knulst, 1982. Mediagebruik bij verruiming van het aanbod, W.R.R. Voorstudies en achtergronden mediabeleid M4, 's-Gravenhage.

Tabel 1. Tijdsbesteding aan enige hoofdcategorieën, uitgedrukt in procenten van de gehele week (168 uur = 100%). Totale steekproef 1975: 1309/1980: 2730

	Arbeid		Huishouding, gezin, boodschappen		Slapen, eten en persoonlijke verzorging		Onderwijs, ontwikkeling		Vrije-tijdsbezigheden		Overige	
	1975	1980	1975	1980	1975	1980	1975	1980	1975	1980	1975	1980
Mannen	14,9	13,0	5,1	5,6	44,7	45,4	4,7	5,0	28,2	28,9	1,7	2,1
Vrouwen	3,2	3,8	19,1	18,9	45,3	45,4	3,3	3,2	26,9	26,2	2,1	2,4
Personen onder 35 jaar	8,5	8,5	12,6	12,3	44,4	44,9	6,5	6,7	26,3	25,3	1,7	2,3
Personen boven 50 jaar	5,3	3,8	13,2	12,9	47,1	47,3	0,3	0,5	31,9	33,1	2,2	2,4
Werkende mannen[1]	26,0	25,2	3,7	4,5	42,7	43,0	0,9	1,1	25,4	24,7	1,3	1,5
Werkende vrouwen[1]	21,7	21,9	9,5	9,3	43,7	43,9	1,1	1,1	22,5	22,2	1,4	1,6
Huisvrouwen	0,8	0,6	23,1	23,5	45,2	45,3	0,6	0,7	28,1	27,3	2,2	2,4
Gepensioneerde mannen	1,3	0,8	10,2	8,5	48,3	48,9	0,2	0,6	37,6	38,8	2,5	2,4
Studenten, scholieren	1,4	1,4	3,9	3,9	46,5	46,6	21,6	20,5	25,2	25,0	1,4	2,6
Personen met lager onderwijs[2]	7,4	5,6	14,4	15,3	46,2	46,7	0,8	0,5	28,7	29,5	2,5	2,3
Personen met middelbaar onderwijs en hoger[2]	10,6	9,3	11,8	10,0	44,0	44,1	3,9	7,1	28,1	27,3	1,6	2,2
Laag beroepsmilieu[3]	8,4	7,2	13,1	14,2	45,2	45,8	3,5	3,4	27,6	27,0	2,1	2,4
Hoog beroepsmilieu[3]	8,1	7,9	14,1	13,1	44,5	45,3	3,9	4,5	27,5	27,0	1,9	2,2
Steekproef gemiddelde	8,2	7,7	13,4	13,3	45,0	45,4	3,8	4,0	27,8	27,3	1,9	2,3

1. Personen die tenminste 20 uur of langer hebben gewerkt.
2. Personen van 18 jaar en ouder.
3. Beroepsmilieu van hoofd van de huishouding.

Tabel 2. Tijdsbesteding aan mediagebruik [1] in uren per week; in procenten van het gehele weekbudget (= 168 uur) en in procenten van de vrije tijd

	Mediagebruik in uren per week		Mediagebruik in % van week		Mediagebruik in % van de vrije tijd	
	1975	1980	1975	1980	1975	1980
Mannen	19,0	20,1	11,5	12,0	40,8	41,5
Vrouwen	15,9	14,7	9,5	8,8	35,3	33,6
Personen jonger dan 35 jaar	16,6	14,9	9,3	8,9	35,4	35,2
Personen ouder dan 50 jaar	22,8	22,3	13,6	13,3	42,6	40,2
Werkende mannen	16,8	16,6	10,0	9,6	39,4	38,9
Werkende vrouwen	10,9	12,0	6,5	7,2	28,9	32,4
Huisvrouwen	16,9	15,6	10,0	9,2	35,6	33,7
Gepensioneerde mannen	27,1	30,1	16,1	17,9	42,8	46,1
Studenten en scholieren	16,1	16,6	9,6	9,9	38,1	39,6
Personen met lager onderwijs	19,8	20,0	11,7	11,9	40,8	40,3
Personen met middelbaar onderwijs en hoger	15,1	16,0	8,9	9,5	31,7	34,8
Laag beroepsmilieu	18,4	17,7	10,9	10,5	39,5	38,9
Hoog beroepsmilieu	15,8	16,3	9,4	9,7	34,2	35,9
Steekproefgemiddelde	17,4	17,0	10,3	10,1	37,1	37,0

1. Mediagebruik: televisie+radio+muziekplaten/geluidsbanden+kranten+tijdschriften+boeken.

Tabel 3. Tijdsgebruik voor verschillende media in 1975 en 1980 in uren per week, gemiddeld over de hele steekproef

	1975 (n = 1309)	1980 (n = 2730)
Televisiekijken	9,8	9,9
Radioluisteren	1,2	1,0
Kranten lezen	2,3	2,2
Tijdschriften lezen	1,9	1,8
Boeken lezen	1,6	1,6
Muziekplaten/banden spelen	0,7	0,6
Totaal media	17,4[1]	17,0[1]

1. Verschil van 0,1 uur in optelling is ontstaan door de tijdscores die gemeten zijn in kwartieren, om te zetten in uren.
2. Als primaire activiteit.

Tabel 4. Verschillen in tijdsgebruik voor media in 1975 naar geslacht, leeftijd en opleidingsniveau en verschuivingen in de verschillen tussen 1975 en 1980

De verschillen 1975-1980 zijn getoetst op basis van de difference- of meanstest, waarbij:
0 = geen significant verschil
+ of − = verschil 1975-1980: 0.05 <p>0.01
+ + of − − = verschil 1975-1980: p ⩽0.01

	Televisie		Radio		Muziekplaten/ banden		Kranten		Tijdschriften		Boeken	
	Index 1975	Verschil 1980 t.o.v. '75	Index 1975	Verschil 1980 t.o.v. '75	Index 1975	Verschil 1980 t.o.v. '75	Index 1975	Verschil 1980 t.o.v. '75	Index 1975	Verschil 1980 t.o.v. '75	Index 1975	Verschil 1980 t.o.v. '75
Gehele steekproef	9,8 uur = 100	0	1,2 uur = 100	0	0,7 uur = 100	0	2,3 uur = 100	0	1,9 uur = 100	0	1,6 uur = 100	0
Sexe												
Mannen	111	+ +	143	0	137	0	124	0	91	0	84	+
Vrouwen	92	0	67	0	74	−	83	− −	108	− −	113	−
Leeftijd												
12 tot 24	91	+ +	93	0	174	− −	49	−	81	− −	118	0
25 tot 49	94	0	60	0	74	0	102	0	96	0	91	0
50 en ouder	126	0	196	− −	70	0	156	0	135	−	105	+
Onderwijsniveau[1]												
Laag	119	+ +	146	0	93	− −	109	0	106	+	86	0
Uitgebreid laag	103	0	83	0	81	−	100	− −	104	− −	94	− −
Middelbaar en hoger	73	+ +	63	0	63	+ +	134	− −	112	− −	116	0

1. van personen van 18 jaar en ouder

Tabel 5. Tijdsbesteding aan media en aantal uren arbeid per week (oktober 1980)

	Uren per week:							
Gemiddeld aantal gewerkte uren per week	Televisie		Radio/Muziek-platen-banden		Kranten, boeken tijdschriften		Media totaal	
	mannen	vrouwen	mannen	vrouwen	mannen	vrouwen	mannen	vrouwen
Geen werk buitenshuis	15,6	9,3	3,3	1,0	9,5	5,6	28,4	15,9
1 tot en met 10 uur	13,5	8,1	4,3	0,9	6,6	5,0	24,4	14,0
11 tot en met 20 uur	10,2	8,5	2,3	0,8	6,9	4,3	19,4	13,6
21 tot en met 30 uur	9,3	6,1	1,6	0,8	6,7	5,2	17,6	12,1
31 tot en met 40 uur	10,3	6,9	1,4	1,1	5,0	4,3	16,7	12,3
41 tot en met 50 uur	9,9	6,7	1,7	1,0	5,1	3,9	16,7	11,6
51 uur en meer	8,4	5,0	1,4	1,5	4,0	3,7	13,8	10,2

Tabel 6. Tijdsbesteding aan media en aantal uren vrije tijd per week (oktober 1980)

	Televisie		Radio/Muziek-platen-banden		Kranten, boeken tijdschriften		Media totaal	
Uren vrije tijd per week	uren	%	uren	%	uren	%	uren	%
minder dan 21 uur	4,3	26,3	0,3	2,0	2,0	11,9	6,6	40,2
21 tot en met 30 uur	7,0	25,9	1,0	2,6	3,2	11,7	11,1	41,2
31 tot en met 40 uur	8,0	23,0	1,1	2,9	4,2	11,8	13,6	37,7
41 tot en met 50 uur	9,8	21,5	1,5	3,3	5,2	11,5	16,5	36,3
51 tot en met 60 uur	10,9	19,7	1,6	3,0	6,2	11,3	18,7	34,0
61 tot en met 70 uur	12,9	20,0	2,7	4,1	7,7	11,8	23,3	35,9
71 uur en meer	14,6	18,4	3,8	4,7	10,0	13,9	29,4	37,0

Tabel 7. Verscheidenheid van vrijetijdsactiviteiten en mediagebruik in één week (oktober 1980)

	Verscheidenheid:[1]			
Media	Gehele steekproef ($n = 2730$)	Lage score (0-5)	Middenscore (6-7)	Hoge score (8 en meer)
Televisie ($r = -.23$)[2]	9,9	11,5 uur	9,8 uur	8,4 uur
Radio ($r = .08$)[2]	1,0	1,3 uur	0,8 uur	0,8 uur
Muziekplaten/-banden ($r = .01$)	0,6	0,6 uur	0,6 uur	0,7 uur
Kranten ($r = .03$)	2,2	2,1 uur	2,2 uur	2,2 uur
Tijdschriften ($r = .02$)	1,8	1,7 uur	1,7 uur	1,8 uur
Boeken ($r = .12$)[2]	1,6	1,5 uur	1,6 uur	1,7 uur

1. Geteld zijn: aantal verschillende soorten vrijetijdsbesteding per persoon in één week, waarbij televisie, radio, kranten en tijdschriften door hun algemeen karakter niet, en boekenlezen, en muziekplaten/-banden draaien wel als zelfstandige vormen van vrijetijdsbesteding zijn beschouwd. Teneinde echter de relatie zo zuiver mogelijk vast te stellen, is bij de diversiteitscore door boekenlezers, het boekenlezen en bij de score van muziekbeluisteraars – het platen draaien – niet meegeteld. De scores varieerden tussen 0 en 14. Door de scores 0-5; 6-7 en 8 en hoger samen te voegen, ontstonden 3 groepen.
2. Produkt-moment-correlatie van mediagebruikscore en verscheidenheidscore. $p < 0.05$.

Tabel 8. Mate van televisiegebruik en gebruik van en tijdsbesteding aan andere media, tijdens één week (oktober 1980)

Mate van televisiegebruik	Gebruik van: Radio		Muziekplaten/ -banden		Kranten		Tijdschriften		Boeken	
	Gebruik	Tijd Index[1]	Gebruik	Tijd Index[1]	Gebruik	Tijd Index[1]	Gebruik	Tijd Index[1]	Gebruik	Tijd Index[1]
Lichte televisiekijkers: minder dan 6 uur per week	29%	67	34%	87	78%	89	81%	92	48%	95
Middelmatige televisiekijkers: 6 tot 12 uur per week	41%	85	34%	95	79%	96	83%	102	52%	93
Zware televisiekijkers: meer dan 12 uur per week	57%	131	33%	121	75%	122	78%	105	45%	115

1. Tijdsbesteding is geïndexeerd aan het gemiddelde van de gebruikers van het desbetreffende medium.

Tabel 9. Mate van televisiegebruik en gebruik van en tijdsbesteding aan verschillende soorten tijdschriften (oktober 1980)

Mate van televisiegebruik	Vrouwenbladen		Populaire (gossip)bladen		Omroepbladen		Hobbytijdschriften		Opinieweekbladen	
	gebr.	kwartieren index	gebr.	kwartieren index	gebr.	kwartieren index	gebr.	kwartieren index	gebr.	kwartieren index
Lichte televisiekijkers (minder dan 6 uur p.w.)	36%	95	15%	79	23%	85	17%	92	13%	98
Middelmatige televisiekijkers (6 tot 12 uur p.w.)	31%	105	25%	92	33%	95	17%	110	12%	100
Zware televisiekijkers	24%	95	26%	124	36%	120	11%	95	8%	104
Totale steekproef	30%	4.1 = 100	22%	3.8 = 100	31%	2.0 = 100	15%	3.9 = 100	11%	5.1 = 100

Tabel 10. Mate van televisiegebruik en tijdsbesteding aan andere media 1975-1980.

0 = geen significant verschil 1975-1980
+ of − = significant verschil $0.05 < p > 0.01$
+ + of − − = significant verschil $p < 0.01$

Andere media		Lichte kijkers (minder dan 6 uur per week)	Middelmatige kijkers (6 tot 12 uur per week)	Zware kijkers (meer dan 12 uur per week)
Radio	Index '75[1]	46	78	176
	Verschuiving '75-'80	0	0	− −
Platen/banden draaien	Index '75[1]	107	89	107
	Verschuiving '75-'80	− −	0	0
Kranten	Index '75[1]	95	103	102
	Verschuiving '75-'80	−	− −	0
Tijdschriften	Index '75[1]	97	103	103
	Verschuiving '75-'80	−	0	0
Boeken	Index '75[1]	105	105	92
	Verschuiving '75-'80	0	0	0

1. Populatiegemiddelde 1975 = 100 (vgl. tabel 4)

6. Ontwikkelingen ten aanzien van aspecten van de progressief – conservatief tegenstelling in Nederland in de jaren '70

Een empirische studie in historisch perspectief*

C.P. MIDDENDORP

1. INLEIDING

Een klassieke sociologische en politicologische probleemstelling is de vraag hoe de maatschappelijke orde tot stand komt en in stand gehouden kan worden. Welke factoren houden het maatschappelijk gebeuren in een althans ogenschijnlijk evenwicht? Hoe blijven de instituties en normeringen in stand die dat soms wankele evenwicht bewerkstelligen? Deze aandacht voor maatschappelijke stabiliteit en orde hangt nauw samen met de ontstaansgeschiedenis van de sociale wetenschappen. De grondleggers van de sociologie opereerden immers in een tijdsgewricht waarin, na de Franse revolutie en de Napoleontische tijd, juist een sterk besef groeide van het bestaan en belang van stabiliserende instituties, normen en waarden, gewoonten en tradities. Het rationele verlichtings-denken was duidelijk tekort geschoten; optimistische ideeën over de rationele en goede mens die kan functioneren als een zelfstandig individu bleken illusies. Het in de praktijk pogen te brengen van maatschappijvormen, gebaseerd op dit soort uitgangspunten, bleek slechts te leiden tot dictatuur en terreur.

De Franse Verlichting vormde wellicht het hoogtepunt van het streven van individuele personen om zich via kritische reflectie op de samenleving los te maken van traditionele, op de Middeleeuwse verhoudingen geïnspireerde kaders. Dit denkproces diende ondersteund te worden door een nieuwe wetenschap: de *'ideologie'* die als studieobject had: het proces waardoor mensen zich ideeën eigen maken over het functioneren van de maatschappij en het totstandkomen en instandhouden van maatschappelijke verhoudingen: iets wat wij thans zouden aanduiden met de term 'politieke socialisatie'[1].

Deze 'ideologie' is als wetenschap nooit van de grond gekomen. Na de schok van de Franse revolutie en de Napoleontische tijd werd het accent radicaal verlegd. Men was niet langer geïnteresseerd in de studie van individuele leerprocessen welke zouden kunnen leiden tot kritische reflectie op en verzet tegen maatschappelijke verhoudingen. Het studieobject van de *sociologie* was juist: die functionele kaders welke de mens blijkbaar nodig heeft – als sociaal wezen – ter oriëntering van zijn gedrag, en welke nodig zijn voor het handhaven van maatschappelijke stabiliteit en orde. Grote onderdelen van de sociolo-

* zie noten pag. 98

gie zoals die zich vanaf het begin van de 19e eeuw ontwikkelden, kunnen worden beschouwd als de wetenschappelijke onderbouwing van wat wel genoemd wordt het 'Romantisch conservatisme': die tegen het Franse Verlichtingsdenken gerichte stroming, welke aan het begin van de 19e eeuw in West-Europa ontstond en volgens Mannheim het meest volledig tot ontplooiing kwam in Duitsland[2].

In tegenstelling tot de Verlichting gaat het conservatisme uit van een *pessimistisch mensbeeld* en van een *positief maatschappijbeeld.* De mens is irrationeel, zwak, tot het slechte geneigd, afhankelijk, egoïstisch, maar ook complex en 'ongelijk', bepaald door erfelijke aanleg. Het zijn juist de maatschappelijke tradities en gewoonten, normen en waarden, instellingen en sociale controlemechanismen welke voor de mens van onmisbaar belang zijn – zij kunnen hem dat vernisje van beschaving bijbrengen waardoor maatschappelijke orde en rust mogelijk worden[3].

Dit conservatieve denken, dat vanuit deze premissen nader kan worden uitgewerkt, zou waarschijnlijk niet van een dergelijk groot belang zijn geworden indien het niet zou zijn ontstaan in een tijdperk waarin nu juist aan deze heilzame maatschappelijke reguleringsmechanismen duchtig werd geknaagd door de opkomende, ontwrichtende werking van de *industriële revolutie.* Volgens Mannheim is het juist een voorwaarde voor het zich bewust worden van het bestaan van tradities en traditionele instellingen, dat deze in de feitelijke ontwikkelingsgang van het maatschappelijk proces worden belaagd en aangetast. Traditionalisme is dan de 'onbewuste' versie van conservatisme: het *bewust* opkomen voor, op basis van het zich realiseren van de waarde van, traditionele instellingen en gewoonten[4].

Heel schematisch kunnen de ontwikkelingen als volgt worden aangeduid. De industriële revolutie veroorzaakte de transformatie van de traditionele standenmaatschappij tot de 'klassieke' klassenmaatschappij – zij het dat die transformatie niet geheel volledig was: elementen van de traditionele maatschappij bleven voortleven in de pas langzamerhand dominant wordende klassenmaatschappij. Adel, geestelijkheid, de boerenstand en de kleine middenstand bleven voor een groot deel bestaan, maar daarnaast ontstonden twee nieuwe klassen: de *industriële bourgeoisie* en het *proletariaat.*

Uiteraard vertoonden deze ontwikkelingen naar hun aard, fasering en onderlinge verhoudingen van de desbetreffende groeperingen, allerlei variaties van land tot land en van streek tot streek, afhankelijk van voorafgaande historische ontwikkelingen in de sociale structuur, geografische ligging, de beschikbaarheid van grondstoffen voor de industrie, onderlinge machtsverhoudingen en processen van staatsvorming welke reeds hadden plaatsgevonden. In dit verband zal ik niet ingaan op deze variaties, maar een poging wagen het grondpatroon te schetsen[5].

In de klassenmaatschappij, nadat sommigen zich bewust waren geworden van de bedreigde traditionele structuren, ontstonden verschillende ideologische stromingen, geënt op de maatschappelijke positie van (a) diegenen die nog geworteld waren in traditionele structuren en/of bij het voorkomen van verdere aantasting daarvan een zeker belang hadden, en (b) de nieuwe groeperingen, welke hun politieke belangen wilden verdedigen door legitimering van hun positie, of juist via radicale hervormingen hun als onrechtvaardig ervaren

positie wilden veranderen. De onder (a) bedoelde groepen werden de 'dragers' van het *conservatisme*: adel, geestelijkheid, boeren en kleine middenstand. Van de onder (b) bedoelde groepen werd het *klassieke economische liberalisme* vooral aangehangen door de nieuwe industriële bourgeoisie en, zij het veel later, het *Marxisme* door diegenen die opkomen voor de belangen van het proletariaat, dat met het voortschrijden van de industriële revolutie gedurende de 19e eeuw vooral in de steden via grootscheepse urbanisatie ontstond.

De *'ideologische constellaties'* die op deze wijze met allerlei variaties in verschillende landen van West-Europa ontstonden, vertoonden veelal een eigen dynamiek en fasering. Aanvankelijk stonden liberalisme en conservatisme tegenover elkaar: conservatieven verfoeiden het liberale individualisme, geënt op materialisme en economische ontwikkeling via het marktmechanisme, hetgeen ontwrichtend werkte op de traditionele maatschappij. Liberalen beschouwden aanvankelijk deze traditionele maatschappij als belemmerend voor de ontplooiing van het vrije individu en voor de economische vooruitgang.

Later, toen de ontwikkelingen onomkeerbaar bleken en het Marxisme met zijn revolutionaire perspectief een veel grotere bedreiging voor de gevestigde maatschappelijke verhoudingen bleek dan het liberalisme ooit was geweest, incorporeerden vele conservatieve denkers het principe van de vrije markteconomie op basis van privé-kapitaal in hun denken, en konden liberalen, nu ze werden geaccepteerd in de bestaande orde, soms ook veel aspecten van de traditionele maatschappij herwaarderen en accepteren. Zo ontstond een *conservatief-liberaal syndroom*, in oppositie tegen allerlei variaties van Marxistische en daarop aanvankelijk geïnspireerde socialistische en sociaaldemocratische stromingen[6]. Heel in het kort kunnen deze ontwikkelingen en hun implicaties voor de sociale wetenschappen als volgt worden geschetst.

a. Het ontstaan van ideologische tegenstellingen in West-Europa kan teruggevoerd worden tot de inbreuk welke de Franse revolutie en de industriële revolutie maakten op de traditionele maatschappij. De maatschappelijke veranderingen die werden veroorzaakt door beide revoluties, brachten een conservatieve reactie teweeg als bewustwording van de dreigende teloorgang van waardevol geachte tradities. Tegelijkertijd ontstonden op nieuwe maatschappelijke groeperingen geënte ideologieën (liberalisme en socialisme), welke echter weinig perspectief boden voor wat betreft hun inhoudelijke denkbeelden over mens en maatschappij: de idealen van deze stromingen bleven steeds vaag, formeel en/of utopistisch.

b. Het ontstaan van de sociale wetenschappen is nauw verbonden met vraagstukken van maatschappelijke stabiliteit en orde, in het kielzog van de conservatieve bewustwording en als reactie op de desintegrerende werking van vooral de industriële revolutie. Vandaar dat relatief weinig aandacht is besteed aan de systematische bestudering van maatschappelijke veranderingsprocessen, zowel structureel als cultureel. Dit werkt door tot op de huidige dag: een gebrek aan theorievorming omtrent processen van maatschappelijke ontwikkeling.

c. Dit laatste werd nog versterkt door de omstandigheid dat bij het tot volle wasdom komen van de sociale wetenschappen als empirische wetenschappen, het veel gemakkelijker was om te trachten statische situaties theoretisch te interpreteren en te toetsen, dan om te proberen sociale processen te traceren en te verklaren. Dit lag ook meer voor de hand in de periode na de tweede wereldoorlog, waarin men opnieuw meer aandacht had voor problemen van orde en stabiliteit dan voor die van (mogelijke, potentiële) verandering. Deze veranderingsprocessen vonden uiteraard wel plaats (geen maatschappij is statisch), maar kwamen pas in de loop van de jaren '60 in alle hevigheid naar voren.

d. Hoewel culturele factoren en patronen een eigen dynamiek vertonen en een effect (kunnen) hebben op structurele maatschappelijke verhoudingen en processen, kan toch, in het kielzog van de industriële revolutie, in post-industriële maatschappijen het wellicht dominerende belang van de technologische ontwikkeling en haar uitstralingseffect op vele gebieden van het maatschappelijk leven zoals hierboven geschetst, moeilijk worden omzeild. Dit werd in Nederland en in vele andere landen onder meer duidelijk in de jaren '60, toen door de technologische ontwikkelingen na de tweede wereldoorlog de welvaart zeer sterk toenam, mede waardoor processen van schaalvergroting en intensivering van de communicatiepatronen op massaniveau (ondermeer door de televisie) tot stand kwamen. Dit gaf aanleiding, zo kan hypothetisch worden gesteld, tot het ontstaan van processen van politisering, democratisering en individualisering. Het cultuurpatroon veranderde tamelijk ingrijpend, vooral op het gebied van huwelijk, gezin, sexualiteit en religie[7]. De *politieke* attituden op massaniveau duiden, voor zover de beschikbare gegevens reiken, op toen nog tamelijk grote stabiliteit, hoewel op elite- en semi-eliteniveau sprake was van turbulentie; allerlei nieuwe politieke partijen en stromingen gingen een belangrijke rol spelen[8].

De korte historische schets welke hierboven werd gegeven van het ontstaan van ideologische stromingen in de 19e eeuw (conservatisme en liberalisme in de meeste gevallen tegenover allerlei vormen van socialisme) èn het ontstaan van de sociale wetenschappen welke voornamelijk waren gepreoccupeerd met het 'verklaren' van de maatschappelijke orde, dienen beide tot uitgangspunt van de studies waarvan in dit artikel een deel wordt gerapporteerd.

Aan het eind van de jaren '60 was de progressief-conservatief 'tegenstelling' opnieuw zeer actueel. Wellicht juist doordat, naar analogie met de 19e eeuw, de maatschappelijke verhoudingen zich ingrijpend leken te wijzigen, was een ideologische reflectie op de ideeën welke ten grondslag kunnen liggen aan dergelijke veranderingsprocessen interessant. De systematische empirische studie van deze verschijnselen nam een aanvang in 1970[9]. In twee daaropvolgende replicaties van deze basisstudie[10] werd vervolgens gepoogd om recht te doen aan de meer dynamische aspecten van maatschappelijke ontwikkelingen, vanuit de overweging dat niet volstaan kan worden met statische momentopnamen. Ook vanuit de wetenschapsfilosofie kan immers met recht worden betoogd dat éénmalige bevindingen slechts een beperkte geldigheid hebben. Het gaat er immers om, ook *in de tijd* te kunnen vaststellen welke generaliseerbaarheid geldt voor bepaalde structuren en relaties, alsmede welke tendenties zich voor-

doen in de cultuur van een maatschappij[11]. Dat men zich hierbij, zoals boven aangegeven, theoretisch op glad ijs dan wel in een vacuüm bevindt, waardoor vele resultaten slechts exploratief kunnen worden geboekt en speculatief kunnen worden geduid, is een nadeel en voor sommigen een belemmering om dit soort onderzoek uit te voeren. Wellicht mede hierdoor, en door het tot nu toe ontbreken van een methodologie van replicerend onderzoek, is het vooralsnog een tamelijk grote zeldzaamheid als breed opgezette studies, gebaseerd op grote representatieve steekproeven uit een bepaalde bevolking, meerdere malen worden uitgevoerd (met bepaalde tijdstintervallen), waardoor trendmatig ontwikkelingen binnen de desbetreffende samenleving op grote schaal zichtbaar kunnen worden[12].

Voor de hierboven bedoelde serie van drie studies (enquêtes gehouden met representatieve steekproeven onder de Nederlandse bevolking in de jaren 1970, 1975 en 1980) geldt bovendien niet slechts dat zij zijn opgezet volgens de 'replikatie-strategie'[13], maar tevens dat zij een streng gedefinieerd begrippenkader als uitgangspunt hanteren: de *progressief-conservatief tegenstelling*. De historische legitimering van deze thematiek is hierboven aangegeven in termen van processen van maatschappelijke ontwikkeling en ideologische, daarop gebaseerde stromingen. Het gaat hierbij om de 'kern' van de politieke cultuur, zoals die zich in West-Europa heeft ontwikkeld in de 19e en 20e eeuw.

Hieronder zal allereerst worden aangegeven hoe deze 'tegenstelling' is gedefinieerd en op verschillende niveaus van 'abstractie' is geoperationaliseerd. Enkele 'aanpalende' bekende verschijnselen zoals anomia, dogmatisme en autoritarisme zullen eveneens worden onderzocht[14]. Het longitudinale aspect van de serie studies doorbreekt het traditionele statische karakter van veel sociaal-wetenschappelijk onderzoek, waarvan de achtergrond, eveneens liggend in de ontstaansgeschiedenis van de sociale wetenschappen in de 19e eeuw, hierboven is aangeduid. De jaren '70, welke de studies bestrijken, lijken voor Nederland, evenals voor vele andere landen, gekenmerkt te worden door een overheersing van maatschappelijke en culturele ontwikkelingen door die van economie en technologie. De jaren '70 bieden het unieke schouwspel van stagnerende welvaart, waardoor grote problemen ontstaan t.a.v. het handhaven van de voornamelijk in de jaren '60 opgebouwde verzorgingsstaat.[15] De effecten daarvan op de ideologische kern van de politieke cultuur zullen hieronder worden beschreven en voor zover mogelijk geïnterpreteerd.

2. METHODOLOGIE

De progressief-conservatief 'tegenstelling' (de vraag of dit begrippenpaar ook *empirisch* een tegenstelling vormt binnen de Nederlandse bevolking komt hieronder aan de orde) is gedefinieerd aan de hand van een *ideaal-typisch model* van de conservatieve ideologie. Immers, zoals in de inleiding is aangegeven, hebben zowel liberalisme als socialisme zich successievelijk als progressieve stromingen tegenover het conservatisme aangediend, waarbij het liberalisme zich in veel gevallen in latere fasen met conservatieve stromingen heeft verbonden. Het progressieve alternatief bood dus weinig houvast, terwijl de conservatieve pool zich gedurende de 19e eeuw in vele opzichten met de oorspronkelijke liberale rivaal verzoende.

Het ideaal-typische model was noodzakelijk gezien de grote diversiteit aan conservatieve en conservatief-liberale denkbeelden, welke zich in West-Europa gedurende de 19e en 20e eeuw in verschillende landen hebben ontwikkeld. Om in deze 'chaos' enige orde te scheppen was het onvermijdelijk om te pogen de essentiële onderdelen van het conservatieve denken samen te brengen en aan elkaar te relateren in een 'model', dat dan opgevat kan worden als een 'theoretische constructie'.

Dit 'model', dat hier uiteraard niet volledig kan of behoeft te worden gepresenteerd, diende als uitgangspunt voor de operationalisatie van de progressief-conservatief tegenstelling. Hier stond de onderzoeker opnieuw voor een dilemma. Het model van het conservatisme bevat immers een aantal gezichtspunten ten aanzien van mens en maatschappij, hun onderlinge relatie, opvattingen over sociale verandering, over maatschappelijke instellingen zoals eigendom, klassen, gezag, democratie en de overheid, die tamelijk *abstract* zijn. *Directe* operationalisatie daarvan zou uitspraken opleveren welke eveneens tamelijk abstract zouden zijn en wellicht niet geschikt voor een breed publiek. Studies waarin dit type items wordt gehanteerd zijn uiterst zeldzaam. Zij leveren echter wel vaak andere resultaten op dan studies waarin de vraagstelling wordt geconcretiseerd en uitmondt in attitudeschalen[16].

De notie van 'ideologie', mede aan de hand waarvan de definitie van de progressief-conservatief tegenstelling plaats vond, bood hier echter uitkomst. Het moest immers mogelijk zijn om één of meer *fundamentele waarden* op te sporen welke ten grondslag liggen aan een dergelijke ideologie, en welke als uitgangspunt konden dienen voor de selectie van meer specifieke attitude 'statements' als onderdeel van attitudeschalen[17]. Aldus werd besloten om het abstracte begrip 'progressief-conservatief' zowel *direct* te operationaliseren op 'filosofisch' niveau (d.m.v. bipolaire uitspraken, waarin zowel het conservatieve als het progressieve gezichtspunt werden verwoord) als *'indirect'* te operationaliseren via een selectie van attitude 'statements' op grond van hun relatie tot de onderliggende waarden van het progressief-conservatief domein: *vrijheid* en *gelijkheid* en hun toepassing op sociaal-economisch en niet-sociaal-economische terreinen. (Zie Middendorp, 1978: pp. 148-151.)

Het ideaal-typische model van het conservatisme dat de progressief-conservatief tegenstelling definieerde, werd in wat genoemd werd de 'directe methode' slechts zeer ten dele geoperationaliseerd: van de 121 modelelementen werden er in de basisstudie in 1970 slechts 37 geoperationaliseerd. In de replikaties in 1975 en 1980 werden er 25 daarvan opnieuw empirisch 'getest'.

Op basis van de waarden *vrijheid* en *gelijkheid* en hun toepassing op het sociaal-economische terrein (met inbegrip van overheidspolitiek op deze gebieden) en het niet-sociaal-economische terrein werden 20 attitudeschalen geconstrueerd op basis van Mokken's H-coëfficiënt, een procedure ontwikkeld op basis van de Guttman-schaaltechniek[19]. Een overzicht van de schalen vindt men in tabel 1; details zijn op aanvraag verkrijgbaar bij de auteur. Tien van deze schalen hebben betrekking op sociaal-economische gelijkheid (tabel 1; schalen 4-13) en negen op niet sociaal-economische vrijheid (schalen (15-23), terwijl de schaal 'Democratie-attitude' t.a.v. inspraak en medezeggenschap betrekking heeft op beide aspecten.

Deze beschrijving van de sets van attitudeschalen loopt vooruit op de be-

Tabel 1. Schaalkenmerken en trends, 1970-1980

Schaalnaam	Range	H-coëfficiënten			Gemiddelde scores		
		1970	1975	1980	1970	1975	1980
1. Conservatisme	0-9				4.50	5.39	5.25
2. Liberalisme	0-9				4.86	5.22	4.92
3. Socialisme	0-9				5.02	5.05	4.90
4. Gelijkheid inkomen, bezit, status	0-4	.49	.51	.45	2.48	2.54	2.57
5. Overh. ingr. t.b.v. gel. inkomen en bezit	0-2	.79	.78	.75	1.15	1.24	1.27
6. Att. t.a.v. sociale wetg.	0-4	.47	.40	.35	1.54	.93	.75
7. Overh. steun onderw.	0-4	.35	.35	.28	2.86	2.60	2.54
8. Ink. pol. overheid	0-3	.46	.46	.30	2.16	2.00	1.85
9. Belastingpolitiek hogere inkomens	0-3	.59	.58	.50	1.07	1.07	1.17
10. Belastingpolitiek lagere inkomens	0-3	.43	.36	.36	1.93	1.79	1.48
11. Dir. econ. ingrijpen overheid	0-3	.41	.41	.31	1.59	1.56	1.66
12. Niet-econ. ingrijpen overheid	0-2	.38	.49	.40	.98	.74	.64
13. Militante vakbondspolitiek	0-2	.54	.45	.44	1.08	.92	.79
14. Democrat. attitude	0-4	.51	.57	.54	2.52	1.82	1.62
15. Vrijheid v. men. uit.	0-6	.41	.50	.47	3.57	3.35	3.78
16. Att. ontwikkel. hulp	0-3	.54	.51	.55	.97	.81	.85
17. Internationalisme	0-4	.42	.33	.49	.95	.89	1.01
18. Tolerantie t.a.v. misd.	0-3	.37	.36	.35	2.09	1.76	1.76
19. Gezinstraditional.	0-5	.42	.34	.33	1.67	1.36	1.41
20. Autoritaire ouder-kind relatie	0-5	.41	.46	.46	2.61	2.36	2.15
21. Conv. man-vrouw rolopvattingen	0-4	.38	.45	.38	1.55	1.19	.84
22. Tolerantie t.a.v. homosexuelen	0-3	.67	.64	.64	2.19	2.28	2.55
23. Premaritale sexuele permissiveness	0-5	.49	.50	.49	3.20	3.32	3.55
24. Orthodoxie[2]	0-8	.68	.65	.64	3.84	3.83	3.74
25. Religieuze toler.	0-3	.78	.82	.74	1.36	1.52	1.51
26. Pol. rel. conserv.	0-4	.38	.41	.43	1.88	1.70	1.52
27. Confessionalisme	0-7	.58	.59	.60	2.67	2.79	2.75
28. Autoritarisme	0-9				5.21	4.97	4.62
29. Dogmatisme	0-9				4.79	4.27	3.87
30. Anomie	0-5	.43	.37	.35	2.07	2.06	1.80
31. LINKS-RECHTS (hoog = links)	0-79				49.39	46.23	45.47
32. LIBERTAIR-TRADITIONEEL	0-107				53.11	53.38	49.26
33. AUTORITARISME[3]	0-76				40.88	41.96	39.30
34. GEZINSTRADITIONALISME[4]	0-24				9.57	8.47	7.86

1. H-coëfficiënten volgens Mokken (1971). Indien geen H-coëfficiënten vermeld: schaal berust op factor-scores.
2. Scores voor schalen 24-27 alleen voor leden van kerkgenootschappen.
3. Niet te verwarren met de korte F-schaal.
4. Niet te verwarren met de schaal van die naam.

schrijving van de analyses welke over de verschillende operationalisaties werden uitgevoerd. In de eerste plaats werden de 25 bipolaire items die de 'directe' (en partiële) operationalisatie van het conceptuele model vormden, onderworpen aan een vergelijkende factoranalyse over de identieke vraagstellingen op de drie tijdstippen. De resultaten daarvan zijn elders gepubliceerd en kunnen hier als volgt kort worden samengevat:

1) De factorstructuur is *relatief stabiel.* In de periode 1970-1975 is zij *zeer* stabiel; in de periode 1975-1980 splitst zich een wat genoemd kan worden 'autoritaire' subdimensie af van de conservatieve hoofddimensie[20].

2) De structuur van de set van bipolaire items vertoont op alle drie tijdstippen drie hoofddimensies:
a. *conservatisme* in 'engere' zin: vragen m.b.t. de waarde van gewoonten en tradities, orde en rust, oude instellingen, het aanbrengen van veranderingen en nog enkele, welke zich in 1980 als afzonderlijke dimensie presenteren.
b. *liberalisme* in de klassiek-economische zin van het vrije ondernemerschap versus overheidsingrijpen;
c. *socialisme* in termen van de onrechtvaardigheid van het bestaan van klassen en standen, het ontstaan van maatschappelijke verschillen door ongelijke kansen, belemmering van vrijheid door bestaande gezagsverhoudingen en de gebrekkige aandacht voor openbare voorzieningen in een maatschappij, gebaseerd op het vrije ondernemerschap.

Het blijkt dus dat, in tegenstelling tot wat geldt voor een politieke elite zoals leden van de Tweede Kamer, de progressief-conservatief controverse als unidimensionele tegenstelling niet bestaat op het niveau van de massa van het electoraat. Theoretische aanzetten ter verklaring van dit verschijnsel zijn weliswaar beschikbaar, maar daarop zal in dit verband niet verder worden ingegaan[21].

Vervolgens werden de 24 sets van items welke gerekend konden worden tot een attitudeschaal, op hun 'longitudinale robuustheid' qua schaalbaarheid getoetst in elk van de drie studies. De resultaten worden samengevat in tabel 1. Zij duiden op relatief grote stabiliteit (dus: robuustheid); slechts een enkele schaal moest vervallen omdat bleek dat door de tijd heen de schaalbaarheid van de vraagstellingen dusdanig verminderde dat de schaal 'ophield te bestaan', en van slechts enkele andere schalen moesten afzonderlijke items vervallen. De vier sets van items met betrekking tot religieuze houdingen vertonen een bijzonder hoge schaalbaarheid en robuustheid.

Vervolgens werd onderzocht in hoeverre de progressief-conservatief controverse 'complementaire' verschijnselen als *autoritarisme, dogmatisme* en *anomie* in termen van factor-ladingen en schaalbaarheid, robuustheid vertoonden in de periode 1970-1975-1980. Een verkorte versie van de 'klassieke' F-schaal en D-schaal vormde een relatief robuuste ongeroteerde één-factorstructuur op de drie tijdstippen, en vijf anomie-items vertoonden tamelijk goede robuuste schaalbaarheid in termen van Mokken's procedure [22].
Ten slotte werden de wat genoemd werd 'filosofische dimensies' van conservatisme, liberalisme en socialisme te zamen met 20 attitude-schalen (met uitzondering van de religieuze attituden) onderworpen aan vergelijkende factoranalyses in de jaren 1970, 1975 en 1980. Ook deze resultaten worden elders gepresenteerd en duiden op zeer grote robuustheid van de structuur van het 'totale' (ook wel genoemd 'ideologische') domein[23]. Er blijken steeds twee hoofddimensies te bestaan en één subdimensie. De hoofddimensies kunnen worden benoemd als *links-rechts in sociaal-economische zin* (bestaande uit de socialisme-dimensie en attitudeschalen m.b.t. gelijkheid en overheidsingrijpen m.b.t. het bevorderen van die gelijkheid in sociaal-economische zin, zoals via het uitbreiden van de sociale wetgeving, het onderwijs, de inkomenspolitiek, de belastingpolitiek, direct economisch ingrijpen, het welzijnsbeleid in beperkte zin en via vakbondspolitiek en democratisering) en als *libertair-*

traditioneel (of, in de 3-factor oplossing: *autoritarisme,* in de klassieke zin van een 'uitgebreide' F-schaal) in termen van de conservatieve filosofie, de F-schaal in termen van de oorspronkelijke items-verkorte versie, de houding t.a.v. democratisering (inspraak, medezeggenschap), politieke attituden zoals t.a.v. vrijheid van meningsuiting, ontwikkelingshulp en internationalisme, en *niet*-politieke attituden zoals ten aanzien van traditionalisme op het gebied van huwelijk, gezin en sexualiteit, de ouder-kind relatie, man-vrouw rolpatronen en tolerantie t.a.v. homosexualiteit en premaritaal sexueel contact.

In de hier niet gerapporteerde analyse blijkt dat in de 3-factor oplossing voor dit 'ideologische' domein de tweede factor weliswaar kan worden geïnterpreteerd als 'autoritarisme' (gezien de hoge ladingen daarop op elk van de drie tijdstippen voor schalen zoals met betrekking tot de klassieke F-schaal, autoritaire ouder-kind relatie, nationalisme, ontwikkelingshulp, conventionele man-vrouw rolopvatting en vrijheid van meningsuiting), maar dat de derde factor een sociale, niet-politieke afsplitsing betreft met vooral hoge ladingen voor de schalen met betrekking tot gezinstraditionalisme en voorhuwelijks geslachtsverkeer. Deze dimensie werd genoemd 'gezinstraditionalisme', in ruimere zin op te vatten dan de specifieke schaal (zie ook noot 23). Op basis van deze analyse van de 2-factor en 3-factor oplossingen van het 'ideologische' domein werden aldus vier gewogen somscores bepaald voor de factoren (1) links-rechts, (2) libertair-traditioneel, (3) autoritarisme en (4) 'gezinstraditionalisme', op basis van wegingsfactoren voor de factorladingen en de standaarddefinitie van de schaalscores (zie Middendorp, 1978: 412 pp. voor details).

De resultaten van de vergelijkende longitudinale analyse in termen van gemiddelde schaalscores op de drie tijdstippen, worden vermeld in de drie rechter kolommen in tabel 1.

3. RESULTATEN[24]

Uit tabel 1 blijkt dat zich in de periode 1970-1975 een tendentie in *rechtse richting* heeft voorgedaan (zie dimensie nr. 31) welke in de periode 1975-1980 sterk is afgezwakt. Het is onduidelijk hoe dit moet worden geïnterpreteerd. Nadere beschouwing van 'deel-trends' ligt in eerste instantie voor de hand. Dan kan het volgende worden geconstateerd.

a. Socialisme is stabiel in de periode 1970-1975 en neemt slechts licht af daarna.
b. Het 'gelijkheidsstreven' (gelijkheid als doelstelling; schalen 4 en 5) is praktisch stabiel.
c. De sterke trend om de sociale wetgeving als 'voldoende' te beschouwen in de periode 1970-1975 zet zich, wat afgezwakt, door in de periode 1975-1980.
d. Het zelfde beeld komt naar voren voor vele andere schalen die deel uitmaken van de links-rechts dimensie, zoals
   - overheidssteun voor het onderwijs;
   - het inkomensbeleid;
   - niet-economisch ingrijpen door de overheid;
   - de houding t.a.v. inspraak (democratische attitude).

e. Andere schalen vertonen een wat afwijkend beeld, zoals
   - belastingpolitiek t.a.v. hogere inkomens: neemt iets toe (in termen van belastingverhoging) in de periode 1975-1980;
   - belastingpolitiek t.a.v. lagere inkomens neemt sterker af (in termen van: niet verlagen) in de periode 1975-1980 dan daarvoor;
   - opvattingen over militante vakbondspolitiek nemen gestaag af.

Hoe kunnen wij deze verschijnselen interpreteren? Blijkbaar komt de bevolking, na een eerste 'ruk naar rechts' als reactie op het zich inzetten van de economische crisis, daarvan langzaam terug. Ofwel: de trend zwakt sterk af. De oorzaken hiervan kunnen de volgende zijn:

a. De bevolking is wel bereid om zijn 'claims' ter zake van overheidsbeleid met een egalitaire inslag te verminderen in tijden van economische malaise, maar aan deze bereidwilligheid zijn grenzen, en deze grenzen worden des te eerder bereikt naarmate (zoals in 1980 het geval was, maar niet in 1975) de regering minder bereid is dan wel in staat blijkt een perspectief te bieden op spoedig herstel van de economie.
b. De rechtse trend in de periode 1970-1975 was een schrikreactie, mede, hoe paradoxaal ook, ingegeven door de wens het de regering gemakkelijker te maken om te regeren in moeilijke omstandigheden door minder hoge eisen te stellen aan het overheidsbeleid.
c. De rechtse trend is een uiting van wantrouwen tegen te ver doorgeschoten overheidsbemoeienis in termen van egalitaire beleidstendenties. In tijden van crisis ontstaat toenemend individualisme ('ieder voor zich'), hetgeen zich uit in weerstand tegen overheidsbemoeienis: zie de tamelijk sterke trend in liberale richting op de liberalisme-dimensie: schaal 2.

Bovenstaande redeneringen ('rationales') sluiten elkaar niet uit, maar welke het 'meest juist' of het minst onjuist is kan niet worden bepaald.

De andere hoofddimensie van ideologische tegenstellingen wordt gevormd door de libertair-traditioneel dimensie of, meer in politieke zin, de autoritarisme-dimensie zoals die is opgebouwd in termen van de korte F-schaal, conservatisme, en een aantal attituden zoals vrijheid van meningsuiting, ontwikkelingshulp, nationalisme en tolerantie t.a.v. misdadigers. Hier treffen wij in eerste instantie het omgekeerde beeld aan vergeleken met de links-rechts dimensie: relatieve stabiliteit in de periode 1970-1975 en een libertaire, anti-autoritaire trend in de periode 1975-1980.

Nadere inspectie van de tabel laat zien dat er in de periode 1970-1975 sprake is van twee tegengestelde tendenties in de componenten van deze dimensie. Op de *politieke* componenten (conservatisme als filosofie, democratische attitude, vrijheid van meningsuiting en ontwikkelingshulp) is er sprake van tendenties in conservatieve of zo men wil 'autoritaire' richting. Op de *'sociale'* componenten (vooral gezinstraditionalisme, autoritaire ouder-kind relatie, conventionele man-vrouw rollen) is er sprake van libertaire tendenties. Opmerkelijk is wel dat 'tolerantie t.a.v. misdadigers' (blijkbaar) als politieke attitude wordt opgevat, maar dat de gemiddelde score op de korte F-schaal iets afneemt en dat de F-schaal dus tot de 'sociale' attituden wordt gerekend. Hoe dit ook zij, de tegengestelde tendenties zijn over het geheel genomen duidelijk

en heffen elkaar globaal praktisch op.

Het is opmerkelijk dat in de periode 1975-1980 de conservatieve trend op de politieke componenten duidelijk wordt teruggebogen of tot staan komt, terwijl de trend in de richting van libertairder opvattingen op niet-politiek gebied zich doorzet en zelfs hier en daar nog wat sterker wordt. Dit is duidelijk zichtbaar bij de conservatisme-dimensie en sterk bij vrijheid van meningsuiting en internationalisme. Democratische attitude neemt weliswaar verder af, doch de trend wordt sterk afgezwakt. Andere politieke componenten zijn stabiel. Daar staat tegenover dat behalve gezinstraditionalisme (component nr. 19) de overige libertaire trends op de 'sociale' componenten zich nauwelijks verzwakt doorzetten en zich nu ook uitstrekken over de componenten 'tolerantie t.a.v. homosexuelen' en 'premaritale sexuele permissiveness'. Ook autoritarisme en dogmatisme nemen overigens verder en tamelijk gelijkmatig af; deze trends worden nu 'ondersteund' door een vermindering van de *anomie*, hetgeen 'longitudinaal' de these ondersteunt dat traditionalisme mede tot stand komt doordat anomie (gevoelens van persoonlijke onzekerheid inhoudend) via persoonlijkheidssyndromen als autoritarisme en dogmatisme leidt tot het zoeken van 'steun' en zekerheid in traditionalistische opvattingen[25]. De trends op de brede libertair-traditioneel dimensie zoals hierboven besproken, komen samengevat ook duidelijk tot uiting op de bredere sets van de componenten autoritarisme (dimensie 33 in de tabel) en gezinstraditionalisme (nr. 34).

Voor de vier religieuze attituden ten slotte blijkt relatief grote stabiliteit voor het gehele decennium. Slechts het politiek-religieus conservatisme (in termen van bv. 'De zonde maakt een betere maatschappij onmogelijk' en 'De bestaande maatschappij stemt overeen met God's wil' neemt gestaag af voor hen die zich beschouwen als lid van een kerkgenootschap. De religieuze tolerantie neemt slechts tussen 1970 en 1975 wat toe en is daarna stabiel. Deze relatieve stabiliteit staat in schrille tegenstelling tot de ontwikkelingen op religieus gebied in de periode 1966-1970 (zie Middendorp, 1974), toen sterke religieus-libertaire tendenties op alle schaalitems optraden. Het is opmerkelijk dat deze stabiliteit in de jaren '70 ook op politiek-religieus gebied (stemmen op het CDA) naar voren komt: sinds de verkiezingen van 1973 hebben de drie christelijke partijen, verenigd in het CDA, nauwelijks verder stemmenverlies geleden.

4. BESPREKING EN BESLUIT

De hier gepresenteerde gegevens vormen de kern en een eerste analyse van een veel bredere set van beschikbare gegevens waarop trendmatige veranderingen in allerlei opvattingen en verwachtingspatronen in de jaren '70 zichtbaar worden. Meer verfijnde analyses door middel van differentiële trends voor subgroepen kunnen aangrijpingspunten bieden voor nadere interpretaties.

De hier gepresenteerde gegevens zijn er een voorproefje van hoe lastig het is om trendmatige ontwikkelingen in een theoretisch kader te plaatsen dat niet voor het overgrote deel speculatief is. En dan hebben wij hier nog te maken met de historisch verankerde ideologische kern van de politieke en aanverwante cultuur, zoals in de inleiding van dit artikel is aangeduid. Bij andere culturele ontwikkelingen is dit wellicht nog lastiger. Daarbij komt het probleem dat

veranderingen in opvattingen, zoals tot uiting komend in de scores op sommige van de attitudeschalen, niet noodzakelijkerwijs 'echte' veranderingen in opvattingen behoeven te indiceren, aangezien de maatschappelijke werkelijkheid waarnaar sommige items verwijzen, zich natuurlijk eveneens kan wijzigen. Hier stuiten wij op de vraag naar de 'mechanismen' welke veranderingen bij individuele personen 'veroorzaken' en het probleem dat de onderhavige sets van data analyses op individueel niveau niet mogelijk maakt. Panel-data, welke dit wel mogelijk zouden maken, zijn echter om praktische en technische redenen moeilijk te verkrijgen over een langere periode.

Het probleem van de 'mechanismen' is waarschijnlijk het meest fundamenteel. Het staat buiten kijf dat deze mechanismen zeer complex zullen zijn, van persoon tot persoon zullen verschillen, vaak onbewust zullen verlopen, en niet los gezien kunnen worden van kwesties van levensfase en veranderingen in sociale positie van de betrokkenen. De massamedia zullen hierbij waarschijnlijk een grote rol spelen. Het lijkt bijvoorbeeld niet onwaarschijnlijk dat de tamelijk grote veranderingen in libertaire zin, welke zich in de tweede helft van de jaren '60 manifesteerden, mede veroorzaakt werden door de snelle toeneming van de massacommunicatie via de televisie, welke toen althans in Nederland op grote schaal zijn intrede deed en via welke veel mensen voor het eerst 'in levenden lijve' en op permanent grote schaal werden geconfronteerd met andere zeden en normen in andere landen[26]. Hoe dit proces verloopt en welke rol 'relevante anderen' daarbij spelen, is echter bijzonder moeilijk te traceren, ook doordat deze processen zoals al opgemerkt, waarschijnlijk ten dele onbewust verlopen. Politieke socialisatie kan men aldus opvatten als een continu proces: eenmaal gevormde meningen behoeven bepaald niet stabiel te zijn; nieuwe informatie en nieuwe omstandigheden en perspectieven kunnen aanleiding geven tot herziening en bijstelling daarvan.

Kan de rechtse en, in politiek opzicht conservatieve of traditionalistische tendens in de eerste helft van de jaren '70 opgevat worden als een eerste 'schrikreactie' van velen op de zich aankondigende en zich al ten dele manifesterende economische crisis, mede in het licht van de onheilspellende suggesties die daaraan vooraf gingen van bv. de Club van Rome? De redenering lijkt niet onplausibel dat deze nieuwe informatie de potentiële politieke eisen van mensen in termen van overheidsbeleid heeft gedempt en velen er wat bewuster van heeft gemaakt dat traditionele verhoudingen, orde en rust hun waarde hebben, wellicht mede doordat de periode welke eraan voorafging nogal turbulent was verlopen. In tegenspraak met deze redeneringen is echter dat in deze periode de anomie niet is toegenomen en dat autoritaire en dogmatische instellingen wat afnemen. Hiermee wordt tegelijkertijd een belangrijk aspect gesignaleerd van een breed opgezet onderzoek: men kan het totale beeld beter en gedifferentieerder overzien en zich hoeden voor plausibel lijkende, doch niet noodzakelijkerwijs juiste redeneringen. Juist in het vlak van het pogen veranderingsprocessen te duiden is dit een bittere noodzaak. Juist hier, niet gehinderd door enige empirische ondersteuning, wordt al te veel orakeltaal serieus genomen.

Echter, de introductie van breed opgezette trend-studies brengt uiteraard zijn eigen problemen met zich mee, bijvoorbeeld die van de stabiliteit van meetinstrumenten, de mogelijkerwijs veranderende betekenis van identieke

meetinstrumenten, de mogelijkerwijs veranderende betekenis van identieke stimuli in de tijd en dus die van de validiteit van de meetinstrumenten. Een andere vraag die zich laat stellen is of veranderingen op (componenten van) ideologische dimensies gerelateerd kunnen worden aan het politieke klimaat in een bepaald tijdvak respectievelijk aan het type 'regime' ofwel: de politieke kleur en samenstelling van het kabinet. Zoals al eerder opgemerkt was er in 1975 voor het eerst sprake van een niet door christen-democraten doch door sociaal-democraten (te zamen met 'gewone' Democraten '66) gedomineerd kabinet. Dit kabinet had zich voorgenomen een actief overheidsbeleid te voeren in gematigd-linkse zin, hetgeen met de zich inzettende economische malaise een tamelijk moeizame opgave bleek. Het feit dat de rechtse tendenties zich vooral sterk deden gelden onder aanhangers van de christen-democratische partijen en de liberalen, en onder de hogere inkomensgroepen[27] zou geïnterpreteerd kunnen worden als 'oppositie' van delen van deze groepen tegen beleid waarin zij slechts weinig heil zagen of dat er mede op gericht zou kunnen zijn hun materiële belangen aan te tasten. Ook hier is het echter weer gevaarlijk om al te snel te besluiten tot de houdbaarheid van dit soort generalisaties. Op een aantal punten is het namelijk zo dat juist de PvdA-aanhangers en de lagere inkomensgroepen de sterkste 'rechtse' trend vertonen, namelijk t.a.v. de belastingverlaging voor lagere inkomens (rechts is tegen) en het voldoende achten van de sociale wetgeving (rechts is voldoende)[28]. Dit zou men dan juist weer kunnen interpreteren in termen van 'solidariteit' met de moeilijke economische situatie waarin het Kabinet Den Uyl moest opereren.

Ten slotte, de ontwikkelingen in de tweede helft van de jaren '70 kunnen mede beïnvloed zijn door zulke politieke factoren als het optreden van het Kabinet Van Agt-Wiegel in 1977 (hoewel de PvdA, ondanks het moeilijke crisisbeleid dat werd gevoerd, bij de verkiezingen van 1977 de grootste overwinning uit zijn geschiedenis boekte) en de voor velen ongeloofwaardige en zwakke oppositie die deze partij tot 1981 tegen dat beleid voerde. Wellicht dat dit wat meer behoudende beleid weer meer ruimte liet voor libertaire neigingen op politiek gebied en de rechtse trends wat heeft doen afzwakken, maar dit is alles speculatie. Het is immers ook mogelijk dat de zich verdiepende economische crisis velen wèl tot meer 'realisme' dwingt en potentiële politieke eisen indamt, maar dat de afzwakking van de eerste trends in de periode 1970-1975 erop wijst dat er grenzen worden bereikt. De tendens in algemene zin tot een terughoudender opstelling van de overheid tegenover het bedrijfsleven (de liberalisme-dimensie) wordt in elk geval bijna geheel teruggedraaid (zie tabel 1, schaal 2). Er is in vele gevallen geen sprake van trends in één bepaalde richting over de gehele, op zich nog beperkte tijdsperiode van één decennium, hetgeen het gevaar van extrapolatie aanduidt alsmede de wenselijkheid om, ondanks alle praktische, technische en theoretische problemen, dit type systematisch trendonderzoek in het kader van een eigen methodologische strategie, over langere perioden voort te zetten. Het vermoeden rijst dat zich pas over *langere* tijdsperiode regelmatigheden zullen gaan aftekenen welke zich lenen voor theoretische interpretatie.

NOTEN

*) Het meest recente materiaal waarop deze bijdrage is gebaseerd (enquête uit 1980) werd verzameld in samenwerking met het Sociaal Cultureel Planbureau van het Ministerie van CRM. Het veldwerk werd uitgevoerd door het NIPO te Amsterdam. Nadere gegevens over de wijze waarop de steekproef is getrokken zijn op aanvraag verkrijgbaar. Over de studies uit 1970 en 1975 werd eerder gepubliceerd: zie Middendorp (1978, 1979) en het Sociaal en Cultureel Rapport (1976, 1978, 1980).

1. Zie bv. Destutt de Tracy, *Eléments d'Idéologie,* Vol. 1,1801.
2. Zie R.A. Nisbet, *Conservatism and sociology,* American Journal of Sociology, 58 (1952): pp. 167-175 en K. Mannheim, Conservative thought, in: P. Kecskemeti (Ed.), *Essays on Sociology and Social Psychology,* London, Routledge and Kegan Paul, 1953: pp. 74-164.
3. Dit wordt in deze studie later behandeld en empirisch verwerkt.
4. Zie Mannheim, 1953.
5. Enkele aanzetten tot een analyse van de variaties in dit proces in de verschillende grotere West-Europese landen en Nederland worden gegeven in mijn nog ongepubliceerde paper 'The development of ideological constellations in Western European Countries'; zie ook Middendorp (1978: pp. 117-126)
6. Voor een nadere uitwerking van dit deel van het betoog kan steeds worden verwezen naar Middendorp (1978: hoofdstuk 4).
7. Voor een aanzet tot theorievorming in deze termen, zie Middendorp (1979: hoofdstukken 3 en 5). Zie ook Middendorp (1974).
8. Deze ontwikkelingen in de jaren '60 mogen als bekend worden verondersteld: de opkomst van de Boerenpartij en D'66, het ontstaan van de PPR en DS'70 als afsplitsingen van de belangrijkste vertegenwoordigers van het naoorlogse 'bestel' KVP resp. PvdA, de Nieuw-links beweging binnen de PvdA, etc.
9. Zie hiervoor weer Middendorp (1978). Dit project werd mogelijk door een subsidie van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek ZWO. Het veldwerk werd uitgevoerd door het toenmalige Bureau Makrotest BV, te Amsterdam.
10. De twee replikaties vonden plaats in 1975 en 1980. De eerste werd mogelijk gemaakt door het Sociaal en Cultureel Planbureau in samenwerking met het Steinmetz-archief, waarvan de auteur toen het hoofd was. Het veldwerk werd uitgevoerd door NIPO BV, te Amsterdam. De tweede replikatie werd wederom uitgevoerd in samenwerking met het Sociaal en Cultureel Planbureau en mede financieel ondersteund door een subsidie van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek ZWO. Het veldwerk was wederom in handen van het NIPO. De methodologie van de replikatie werd uitgewerkt in Middendorp (1979a).
11. Gegeven het dynamische karakter van de werkelijkheid worden door middel van replicerend onderzoek tegelijkertijd twee functies vervuld: (1) de generaliseerbaarheid van éénmalige bevindingen wordt vastgesteld, en (2) maatschappelijke ontwikkelingen kunnen empirisch worden vastgelegd. Dit maakt zelfs simpele 'herhaling' van onderzoek zo aantrekkelijk. Replikatie houdt echter veel meer in dan simpele herhaling; het gaat daarbij om de ontwikkeling en verdere theoretische uitbouw van 'research designs'; zie Middendorp (1979a).
12. Behalve het project waarover hier wordt gerapporteerd zijn er nog enkele projecten waarin trendmatige veranderingen op relatief grote schaal kunnen worden bestudeerd, zoals de Nationale Kiezersonderzoekingen van 1971, 1972-73, 1977 en 1981, de serie van 24 'golven' van het zogenoemde 'Kontinu Onderzoek' van de FSW-A van de Universiteit van Amsterdam sinds 1971, en het project 'Changing Mass Publics'; een internationaal vergelijkend en 'longitudinaal' project waarin de Katholieke Hogeschool van Tilburg participeert (zie bijv. Heunks 1980). Zie ook voor een beschouwing over de functies die dit soort projecten kunnen hebben binnen bv. de politicology: Middendorp (1982).
13. Zie Middendorp (1979a).
14. *Anomia* (ook wel aangegeven als 'anomie' of 'vervreemding') is als begrip nauw gerelateerd aan die 18e en 19e eeuwse ontwikkelingen, in de aanvang van dit artikel geschetst, welke de traditionele orde bedreigden en verstoorden: de Franse en industriële revolutie. Hoewel volstrekt niet duidelijk is of in vroeger tijden mensen niet evenzeer 'vervreemd' konden geraken van traditionele ordenings'mechanismen', werd deze kwestie ogenschijnlijk acuut ten tijde van de 'aantasting' van vertrouwde maatschappelijke kaders gedurende de 19e eeuw. Het begrip 'anomie' werd bekend door de behandeling ervan door één der grondleggers van de so-

ciologie (Durkheim, 'Le Suicide') en keerde als 'aliënatie' of 'vervreemding' terug in de Marxistische benaderingswijze: de arbeiders 'vervreemd' van de natuurlijke produktiewijzen. Hoewel 'anomia' later een moderne betekenis heeft gekregen, onder meer door het werk van Seeman (1959), ligt de oorsprong van het begrip daar waar de eerste sociale wetenschappers zich bewust werden van de desoriëntatie welke voor velen het gevolg kon zijn van de aantasting van de traditionele orde. De latere toepassingen op de 'moderne maatschappij' zijn daarvan afgeleid.

Autoritarisme is een op het fascisme geïnspireerd begrip, waarbij bedacht kan worden dat het fascisme en nationaal-socialisme als extreme 'aberraties van het conservatisme kunnen worden opgevat (zie Middendorp, 1978, p. 132).

Dogmatisme is een begrip dat is geïnspireerd door autoritarisme. Het zou de meer algemene houdingsstructuur ofwel persoonlijkheidsoriëntatie reflecteren van niet slechts 'rechts' doch meer algemeen 'autoritarisme'. Het begrip is geïnspireerd door de koude oorlog en de gedachtengang dat ondemocratische oriëntaties niet slechts kenmerkend zijn voor het fascisme maar ook voor het communisme, waarbij werd verondersteld dat beide een grondslag vinden in een persoonlijkheidsstructuur. Verder kan op deze kwesties in dit verband niet worden ingegaan: zie verder bijvoorbeeld Seeman (1959) en Strole (1956), Adorno et al (1950) en Rokeach (1960).

15. Voor recente beschouwingen over de problematiek van de verzorgingsstaat zie bv. Van Doorn en Schuyt (1978), en het tijdschrift 'Beleid en Maatschappij' (1980; februari).
16. Zie bijvoorbeeld McClosky (1958), Prothro and Grigg (1960) en Free and Cantril (1967).
17. De definitie van ideologie zoals gehanteerd in deze studie, luidde als volgt: een ideologie is 'een systeem van algemene ideeën over mens en maatschappij, gecentreerd rond één of meer centrale waarden, dat zichzelf historisch als een 'doctrine' heeft gemanifesteerd, welke werd aangehangen door bepaalde groepen of sociale categorieën en dat de belangen reflecteert van een bepaalde belangrijke sociale categorie, bijvoorbeeld een sociale klasse'. Hierbij zij nog aangetekend dat het in het algemeen ook *meerdere* sociale groepen kunnen zijn, welke hun belangen uitgedrukt zien in de ideologie, zoals het geval is bij het conservatisme. Voor de analyse in termen van vrijheid en gelijkheid en hun toepassing op de economische en niet-economische aspecten van de sociale werkelijkheid, zie Middendorp (1978: pp. 148-151).
18. In de 1980-replikatie werden veel meer elementen 'direct' geoperationaliseerd, maar aangezien hiervoor in de tijd geen vergelijkingsbasis bestaat wordt dit hier buiten beschouwing gelaten.
19. Zie Mokken (1971). Mokken hanteert een stochastisch model in plaats van het deterministische Guttman-model.
20. Dit betreft vier vragen met betrekking tot: de onmogelijkheid van volledige democratie, de wenselijkheid van de uitoefening van gezag door een kleine elite van bekwamen, de onmogelijkheid de maatschappij te verbeteren en (zwak) de onvermijdelijkheid van bestaande sociale verhoudingen, zie Middendorp (1982a).
21. Voor de leden van de Tweede Kamer werd althans in 1970 een unidimensionele structuur gevonden (zie Middendorp 1978: pp. 203-204). Aanzetten tot een theoretische verklaring kunnen gevonden worden in Kerlinger's (1967) notie van de 'criteriality' van attitudinele referenten, en een generalisatie daarvan tot waarden zoals vrijheid en gelijkheid.
22. De verkorte F- en D-schalen berusten op geselecteerde items uit de klassieke schalen van Adorno et. al (1950) en Rokeach (1960). De anomia-items zijn grotendeels ontleend aan het werk van Seeman (1959).
23. Zie Middendorp (1982a). De vier religieuze attituden 'verstoren' de over-all factoriële oplossing enigszins (voor leden van een kerkgenootschap), maar de resultaten worden daarmee niet oninterpretabel: een sociaal-religieuze factor wordt duidelijk onderscheiden van een politieke autoritarisme factor en de links-rechts factor.
24. Alle details over de schaalitems en de schaalbaarheid daarvan in de tijd zijn verkrijgbaar bij de auteur.
25. Voor deze redenering, zie Middendorp (1978: pp. 320-327).
26. CBS-gegevens t.a.v. de toeneming van het televisiebezit gedurende de jaren '60 spreken hier duidelijke taal.
27. Zie Middendorp (1979: pp. 102-164).
28. Idem pp. 133-135.

BIBLIOGRAFIE

Adorno, F.W., E. Frenkel-Brunswik, D.J. Levinson, R.W. Sanford, 1950. *The Autoritarian Personality,* New York, Harper and Row.

Beleid en Maatschappij, februari 1980. Themanummer over de problematiek van de verzorgingsstaat.

Destutt de Tracy, 1801, *Eléments d'Idéologie,* Vol 1.

Doorn, J.J.A. van, C.M. Schuyt (Red.), 1978. *De stagnerende verzorgingsstaat,* Boom en Zoon, Meppel.

Free, L.A., H. Cantril, 1967. *The political beliefs of Americans,* New Brunswik, Rutgers, University Press.

Heunks, F.J., 1979. *Nederlanders en hun samenleving,* APA (Holland Universiteits Pers), Amsterdam.

Kerlinger, F.N., 1967. *Social attitudes and their criterial referents; a structural theory,* Psychological Review, 74, 2: pp. 110-122.

Mannheim, K., 1953 (1927). *Conservative thought,* in: P. Kecskemeti (Ed.), Essays on Sociology and Social Psychology, London, Routledge and Kegan Paul.

McClosky, H.M., 1958, *Conservatism and Personality,* American Political Science Review, 52, 1:27-45.

Middendorp, C.P., 1974. *Culturele veranderingen in Nederland, 1965-1970,* Intermediair, 10, 11:1-9.

1978. *Progressiveness and conservatism. The fundamental dimensions of ideological controversy and their relationship to social class,* Mouton Publishers, The Hague etc.

1979. *Ontzuiling, politisering en restauratie in Nederland. Progressiviteit en conservatisme in de jaren 60 en 70,* Boom en Zoon, Meppel.

1979a. Replication as a methodology to realise cumulation of knowledge in the social sciences, *Paper presented at the 11th World Congress of the International Political Science Association at Moscow,* August.

1980, The development of ideological constellations in Western European countries, Paper presented at an international Symposion of the Interdisciplinary Studygroup on European Integration (ISEI), The Hague, Amsterdam.

1982, Over het wetenschappelijk belang van 'onderzoekstradities' in de politicologie, Acta Politica, 2 (juli).

1982a, Are established ideologies in a state of crisis? The case of the Netherlands in the 1970's, Paper presented at the 12th World Congress of the International Political Science Association at Rio de Janeiro, August.

Mokken, R.J., *1971. A Theory and Procedure of Scale Analysis, with applications in political research,* Mouton, The Hague.

Nisbet, R.A., 1952. *Conservatism and sociology,* American Journal of Sociology, 58: pp. 167-175.

Prothro, J.W.C.M. Grigg, 1960. *Fundamental principles of democracy: bases of agreement and disagreement,* Journal of Politics, 22: pp. 276-294.

Rokeach, M., 1960. *The open and closed mind; investigations into the nature of belief systems and personality systems,* New York, Basic Books.

Seeman, M., 1959. *On the meaning of alienation,* American Sociological Review, 24: pp. 783-791.

Sociaal en Cultureel Planbureau, 1976, 1978, 1980. *Sociaal en Cultureel Rapport,* Staatsuitgeverij, Den Haag.

Srole, L., 1956. *Social integration and certain corrolaries, an exploratory study,* American Sociological Review, 21: pp. 709-716.

7. Twee wijzen van generaliseren: extensief en intensief onderzoek

C.J.M. SCHUYT

1. INLEIDING

Het marktonderzoek heeft in grote lijnen de ontwikkeling in de sociale wetenschappen gevolgd. Men zou kunnen zeggen dat het marktonderzoek vooral de opkomst van de sociologische onderzoekingen direct na de Tweede Wereldoorlog heeft gevolgd. Deze opkomst was zelf een reactie op de vooroorlogse beoefening van de sociale wetenschappen, waarin vooral in Nederland een sociografische richting op de voorgrond getreden was: Kleine, gedetailleerde studies over duidelijk beperkte leefgemeenschappen (een dorp in Drente, de Wieringermeerpolder, of de proletarische achterhoede, of de textielindustrie in Twente). Als reactie op de *eerste* fase schoot de *tweede* fase van het sociaalwetenschappelijk onderzoek met in haar directe kielzog het marktonderzoek in een sterk op generalisatie naar grote aantallen gerichte richting. Deze richting is een lange periode dominant geweest. Nu lijkt zich in de beoefening van de sociale wetenschappen weer in recente jaren een kentering voor te doen naar onderzoek, dat wederom meer op gedetailleerder waarneming en op kwalitatieve inductie is gebaseerd. Deze slingerbeweging in onderzoekmethode en -stijl kan van belang zijn voor het marktonderzoek. In dit artikel wil ik nagaan in hoeverre het nu langzaam ontwikkelde intensieve onderzoek in de sociaalwetenschappelijke methode zich onderscheidt van het traditionele onderzoek en welke mogelijke consequenties eruit zouden kunnen volgen voor het marktonderzoek. Deze *derde* fase is in feite nu pas in volle gang. Om het belang van deze derde fase aan te geven wil ik toch nog eenmaal terugkeren naar de periode voordat het grootscheepse en grootschalige onderzoek van de tweede fase opkwam.

Tussen de twee wereldoorlogen is er in Nederland een beweging geweest, die er naar streefde om de verschillen in de betekenis die mensen aan woorden toekennen, systematisch in kaart te brengen. Bekende vertegenwoordigers van deze 'signifische' beweging waren Frederik van Eeden, de jurist Jacob Israël de Haan en de wiskundigen Brouwer en Mannoury. Vooral de laatste heeft ervoor geijverd om door middel van onderzoek de spreiding in woordbetekenissen na te gaan.

De sociale wetenschappen alsook het marktonderzoek stonden toen nog in de kinderschoenen, zodat Mannoury ook op dit gebied enigszins pionierde, slechts geholpen door zijn logische en wiskundige geest en instelling. Hij suggereerde twee methoden om de variatie in woordbetekenissen te onderzoeken en experimenteel vast te stellen: de eerste methode bestond erin om naar *zo-*

veel mogelijk verschillende situaties te vragen, waarop *één en dezelfde persoon* verbaal reageert, de andere om de verbale reacties van *zoveel mogelijk en zo verschillend mogelijk georiënteerde personen* te registreren op *één en dezelfde situatie.* Een van de resultaten van dit onderzoek naar woordbetekenissen was dat Mannoury constateerde dat vele begrippen een aanzienlijk geringere intersubjectieve draagwijdte hadden dan gewoonlijk werd aangenomen.

Het sociaal-wetenschappelijk onderzoek na de Tweede Wereldoorlog heeft in haar ontwikkeling voor een soortgelijke keuze gestaan en heeft onder verschillende invloeden vooral gekozen in de richting van het type onderzoek, waarbij véél personen onderzocht werden in meestal één min of meer dezelfde situatie. Drie vooronderstellingen lagen hieraan bewust of onbewust ten grondslag, namelijk 1) dat de individuele persoon het voornaamste object van onderzoek zou dienen te zijn (de onderzoekseenheid), 2) dat objectieve waarneming van de werkelijkheid mogelijk was en dat onderzoek(st)ers zo weinig mogelijk subjectieve invloed dienden uit te oefenen op de registratie van de onderzochte werkelijkheid (vandaar een streven naar objectief registreerbare waarnemingstechnieken, zoals de pencil-and-paper techniek en het voorgestructureerde interview), 3) dat er een grote mate van eenduidigheid was aan te nemen bij de betekenissen, die ondervraagde personen aan woorden en aan woordcombinaties toekenden.

In deze tweede fase onderzocht men voornamelijk registreerbare kenmerken van (proef)personen en de statistische relaties tussen deze kenmerken. Het proces van communicatie tijdens het interview, de eenduidigheid van woordbetekenissen en de invloed, die interview(st)ers onbewust hadden op het gehele proces van kennisverwerving, werd zelf niet aan nader onderzoek onderworpen. In de optimistische verwachting dat de sociale wetenschappen in principe algemene wetmatigheden van het sociale leven zouden kunnen vinden, werd het merendeel van het sociale onderzoek op deze sterk procedurele en bureaucratische leest geschoeid. De komst van nieuw rekentuig in diezelfde periode (1955 – '65) versterkte deze ontwikkeling naar onderzoek van véél personen in min of meer vergelijkbare situaties. De vraag naar dergelijk onderzoek steeg eveneens. Het kon veel gegevens bevatten en deze konden op geordende manier worden gepresenteerd: de kwantificering in het sociale onderzoek op grote schaal. De andere manieren om onderzoek te doen, de uitgebreide participerende observatie of het kwalitatieve of diepte-interview en groepsdiscussies verdwenen weliswaar niet helemaal naar de achtergrond, maar werden toch veel minder gebruikt. Ook de waarschuwing van Mannoury dat de spreiding van betekenissen, die personen aan woorden geven wel zeer groot is, werd misschien niet helemaal vergeten, maar in ieder geval niet serieus als probleem gezien. De logisch alternatieve methode die Mannoury zelf al gegeven had, werd zo goed als vergeten.

In de aldus ontstane situatie in het na-oorlogse sociale onderzoek heeft het marktonderzoek grote toepassingen gekregen. Zo sterk zelfs, dat de verschillen tussen het wetenschappelijke of zuiver wetenschappelijke onderzoek en het snel in speciale marktonderzoekbureaus ondergebrachte toegepaste onderzoek voor de buitenstaanders en de leken soms moeilijk zijn aan te geven. Het 'beeld' dat al of niet geïnteresseerde leken hebben van sociaal-wetenschappelijk onderzoek, is heel vaak ontleend aan wat men aan opinie-onderzoek kent.

Zo veel % van de mensen vindt dit, zo veel % doet dat en dergelijke. Daar is op zich niets op tegen, maar deze situatie bergt het gevaar in zich dat de eigenlijke taak van wetenschappelijk onderzoek uit het oog verloren wordt. De sociologie hoeft niet te vinden (in de zin van uitzoeken, ontdekken) wat de mensen vinden (in de zin van opinies erop na houden), maar zou juist moeten vinden (ontdekken) wat de mensen zelf niet vinden (zelf niet kunnen of willen ontdekken). De sociale wetenschappen moeten niet vinden wat de mensen vinden, maar zouden juist moeten vinden wat de mensen zelf niet vinden.

Soms brengt markt- en opinie-onderzoek inderdaad resultaten aan het licht, die niet door persoonlijke ervaring of inzicht kunnen worden ontdekt (bijvoorbeeld onderzoek naar oneerlijke concurrentie en misleidende reclame, waarbij het persoonlijke inzicht van een rechter wordt gecontroleerd en gecorrigeerd door de uitkomsten van marktonderzoek). Deze situaties zouden zich dus lenen voor verder gebruik in wetenschappelijke zin: waarom gedragen mensen zich zoals ze zich gedragen en waarom kenden de meesten dergelijke gegevens niet? Door de commerciële setting en de primaire gerichtheid op dienstverlening komen marktonderzoekbureaus meestal aan deze waaromvraag niet toe. Hier ligt duidelijk hun beperking, maar deze beperking ontlokt bij mij enerzijds een kritiek op het veelvuldig gebruik van dergelijk opgezette onderzoekingen in bijvoorbeeld het (overheids)beleid, anderzijds een pleidooi om – ook in het marktonderzoek – te komen tot een andere wijze van onderzoek, eventueel in combinatie met het bestaande en traditionele onderzoek. Ik zou deze andere methode niet de 'kwalitatieve' methode willen noemen, maar liever de 'intensieve' methode en wel om de volgende reden. Kwalitatief suggereert een tegenstelling tot kwantitatief, maar de nieuwere onderzoekmethoden hoeven in principe niet zonder getalmatige abstracties (= kwantificering) te werken. Ook intensief onderzoek kan tot tellen, meten en rekenkundige abstracties leiden. Bovendien berust de onderscheiding 'extensief' versus 'intensief' onderzoek op het onderscheid dat in de logica gemaakt wordt tussen de extensie en de intensie van een uitspraak: de extensie slaat op het aantal *leden* van een klasse dat een bepaald kenmerk heeft, de intensie slaat op het aantal *kenmerken* dat aan één lid kan worden toegeschreven.

In principe kunnen we aldus verschillende typen van onderzoek construeren volgens deze twee logische dimensies: 1) onderzoek naar *véél* kenmerken bij een klein aantal personen (logisch minimum is één); 2) onderzoek naar véél kenmerken bij véél personen; 3) onderzoek naar weinig kenmerken bij véél personen; 4) onderzoek naar weinig kenmerken bij weinig personen. Het eerste type wordt het intensieve onderzoek genoemd, het derde type het extensieve. Het tweede type (veel-veel) zou in principe een ideale onderzoeksituatie opleveren, maar stuit in het algemeen op veel praktische bezwaren van uitvoering en organisatie. De vierde mogelijkheid lijkt wetenschappelijk de minste perspectieven te bieden.

In zekere zin zijn we bij de indeling intensief-extensief weer terug bij de zelf ontworpen indeling van Mannoury bij zijn onderzoek naar woordbetekenissen. Zijn suggestie om één en dezelfde persoon in een lange reeks van situaties te volgen is onlangs zelfs – uiteraard zonder de verwijzing naar Mannoury – gevolgd toen een socioloog verslag uitbracht van een onderzoek, waarbij aan één persoon gedurende *tien* jaar gevraagd was *elke* zaterdag een verslagje in te

leveren over de wederwaardigheden van die week. Onderwerp van het onderzoek was... de gevoelens van verliefdheid. Waar de logica ons al niet toe kan brengen.

In het vervolg van dit artikel zal ik in paragraaf 2 eerst de kritiek op het traditionele onderzoek behandelen en de kentering aanwijzen, die zich in bepaalde centra van wetenschapsbeoefening aan het voltrekken is. In de derde paragraaf zal ik komen tot een vergelijking van het extensieve met het intensieve onderzoek en de toepassing ervan in het marktonderzoek. Daar zal blijken dat de verschillen tussen beide vooral neerkomen op de wijze van generaliseren die gebruikt wordt.

2. KRITIEK EN KENTERING

Het standaardtype van het traditionele onderzoek in de sociale wetenschappen en vooral bij het markt- en opinie-onderzoek zou ik als volgt willen beschrijven:

- bij een zo groot mogelijk aantal personen (bij voorkeur niet minder dan 800 en tenminste 400)
- die op een aantal kenmerken (meestal leeftijd, geslacht, geografische locatie en waar mogelijk ook beroep en inkomen) statistisch representatief geacht kunnen worden
- worden meningen en opinies gepeild, voorkeuren of afkeuren geregistreerd, feitelijke gedragingen die verbaal gereconstrueerd kunnen worden, geregistreerd
- door middel van een eenmalig interview van beperkte omvang (meestal niet langer dan 1½ à 2 uur, meestal minder)
- uitgevoerd door speciaal geïnstrueerde interview(st)ers, die erin getraind worden zo objectief mogelijk te registreren (d.w.z. met uitsluiting van eigen interpretaties en met zo veel mogelijk vermijding van dubbelzinnigheden in de antwoorden van respondenten (het zgn. dóórvragen)).

De aldus geregistreerde kenmerken worden op een getalmatige wijze gecodeerd en met elkaar in verband gebracht. De relaties en correlaties tussen de kenmerken vormen de belangrijkste grondslag voor de weergave van... ja, van wat precies? Van de werkelijkheid? Van de verbale reacties op verbale prikkels, die op een gegeven moment aan de respondenten zijn toegediend? Van het koop- of kiesgedrag van de ondervraagde personen? Van de tendenties in de publieke opinie?

Deze vragen zijn uiteraard al vaker gesteld aan de traditionele onderzoekmethode en ik hoef ze hier niet te herhalen. Ook de kennistheoretische premissen die aan deze methoden ten grondslag liggen zal ik niet naar voren halen. Zelfs alle onvolkomenheden bij de registratie van de kenmerken en alle fouten die daarbij kunnen optreden vormen niet mijn voornaamste punt van kritiek. Immers, geen enkele meetmethode, ook niet in de natuurwetenschappen, werkt zonder meetfout. Er zijn bovendien statistische methoden om deze fouten op hun systematiek te controleren.

Ik wil mijn kritiek op dit standaardmodel op twee andere punten richten, met name 1) op de vraag of er op deze wijze ooit een afdoende *verklaring* zal

kunnen worden gevonden voor de te beschrijven kenmerken van gedrag van mensen en 2) op de vraag hoe adequaat de toepassingen zijn, bijvoorbeeld in het beleid, die op dergelijk onderzoek zijn gebaseerd.

Het waarom van de geobserveerde verschijnselen intrigreert de wetenschapsbeoefenaar in hoge mate en is ook voor de marktonderzoeker van uitermate groot belang. Naast het aandeel van persoonlijke nieuwsgierigheid en avontuurlijkheid die in het uitzoeken van het waarom verscholen ligt, komt er een heel praktische kant van de waarom-vraag naar voren. Als precies kan worden beschreven hoe de werkelijkheid er uitziet en waarom die er uitziet zoals ze er uitziet, dan is de praktische verwachting gewettigd, dat ons handelen beter rekening zal kunnen houden met die 'weerbarstige' werkelijkheid. Met andere woorden er is een realistische kant aan het beantwoorden van de waarom-vraag. We kunnen beter met de werkelijkheid rekening houden als we er achter gekomen zijn hoe die werkelijkheid er *precies* uitziet. Maar daar ligt de crux van mijn bezwaar tegen het standaardtype onderzoek. Men neemt niet de moeite en de tijd om *precies* uit te zoeken hoe die werkelijkheid er uitziet maar men volstaat met een sterk gereduceerde kijk op de sociale werkelijkheid. De enkele kenmerken die men bij veel personen registreert blinken uit in oppervlakkigheid. Niet naar het waarom van de persoonlijke voor- of afkeuren wordt uitvoerig ondervraagd, maar de enkele slotsom: er voor, er tegen of weet-niet. Men ontloopt hiermee niet het gevaar dat de interviewer personen naar iets vraagt, waarover de ondervraagde eigenlijk niet een mening heeft. De sociale werkelijkheid is veel subtieler dan men ooit via dit standaardonderzoek zal kunnen vinden. Aan de ambivalenties, die zo kenmerkend zijn voor meningsvorming bij personen, aan de situationele beïnvloeding, die eveneens kenmerkend is, aan de groepsmeningsvorming, de gelaagdheid in de belangrijkheid van meningen, wordt allemaal geen recht gedaan. Dit zal natuurlijk snel beaamd worden door degenen die niettemin voorstander blijven of gebleven zijn van dit type onderzoek. Zij zullen een eenvoudiger argument er tegenover stellen: het werkt. Op grond van de weliswaar sterk vereenvoudigde en sterk gereduceerde blik op de sociale werkelijkheid blijken adequate voorspellingen mogelijk over bijvoorbeeld koop- en kiesgedrag en zelfs over verschuivingen in de opinies over een overzichtelijke periode (bv. verkiezingsonderzoek). In het algemeen gaat het hier om relatief simpele gedragingen. Met andere woorden men hoeft niet het volledige waarom te weten om toch adequate antwoorden te vinden op praktische probleemsituaties. De marge van onzekerheid kan statistisch worden aangegeven en de op dergelijke onderzoekingen gebaseerde verkoopmethoden en reclames blijken te werken. De beschrijving van gedrag op aggregaatniveau is voldoende.

Hier lopen de wegen van zuiver wetenschappelijk onderzoek en toegepast onderzoek uiteen. Dit praktische argument moge weliswaar voor bepaalde doeleinden voldoende zijn, voor de beantwoording van door de wetenschap gestelde vragen mág het eenvoudig niet voldoende zijn. De zuivere wetenschap wil dóórvragen, maar dan op een andere manier als in het standaardinterview. Doorvragen naar de onderliggende verklarende mechanismen, die ineens een veelheid van ogenschijnlijk onsamenhangende gedragingen met elkaar in verband kunnen brengen.

Het voortijdig laten rusten van de waarom-vragen is voor een wetenschap-

pelijk onderzoek(st)er een doodzonde. Om die principiële reden kan men in de sociale wetenschappen niet volstaan met het gestandaardiseerde onderzoek, dat de werkelijkheid op tamelijk oppervlakkige wijze benadert.

In dit verband is het nuttig op te merken dat de grote doorbraken in biologie en chemie niet zijn voortgekomen uit het grootschalig met elkaar correleren van zoveel mogelijk kenmerken van bestudeerde verschijnselen, maar juist uit het smeden van *nieuwe* en verfijnde waarnemingstechnieken (eerst de microscoop en vervolgens de elektronenmicroscoop). Door de mogelijkheid van *gedetailleerde waarneming* bleken ineens veel meer verschijnselen met elkaar in verband te brengen te zijn. De analogie met deze ontwikkeling in de scheikunde, ook al gaat deze nooit helemáál op, is van betekenis voor de sociale wetenschappen. In plaats van te wachten op de Newton van de sociale wetenschappen op zoek naar mechanische verbanden tussen geaggregeerde grootheden zouden de sociale wetenschappen ook kunnen streven naar hun Virchow of Pasteur. Dit vereist echter *nauwkeurige* en *langdurige* observaties van beperkte objecten of eenheden. De waarnemingstechnieken, die overheersend gebruikt worden in de standaardonderzoeken van de sociale wetenschappen zijn juist gebaseerd op ongeduld en globaliteit. Ze zijn derhalve bruikbaar waar de uitkomsten globaal kunnen of mogen zijn: marktonderzoek en opinie-onderzoek en sommige soorten sociaal-wetenschappelijk onderzoek.

De kentering in de sociale wetenschappen ten gunste van arbeidsintensievere onderzoekmethoden, van langere duur en van grote complexiteit, heeft dus goede wetenschappelijke achtergronden. Ook aan de rol van taal en woordbetekenissen kan in dit nieuwe onderzoektype meer aandacht besteed worden. Naast kentheoretische argumenten voor een nieuwe weg in de methoden van onderzoek zijn er dus ook wetenschappelijke argumenten te formuleren: door betere en langduriger waarneming en waarnemingstechnieken is eerder een verklaring te verwachten van de bestudeerde verschijnselen dan door stil te staan bij de beproefde, maar oppervlakkig blijvende, standaardmethoden. Dat deze kentering voor het markt- en opinie-onderzoek problemen oplevert is duidelijk. Dit heeft immers om praktische redenen minder behoefte aan de beste wetenschappelijke theorie of verklaring en bovendien is de gehele organisatie van het onderzoek juist gericht op de kortstondige waarneming. Het is zelfs de vraag of de intensievere methode van onderzoek wel commercieel te exploiteren valt.

Toch is het ook van belang het andere argument tegen de standaardonderzoekmethode nader te beschouwen. Als het zo is dat er slechts een oppervlakkige benadering van de werkelijkheid mee bereikt kan worden, dan is het logisch om te veronderstellen dat de *toepassingen* in de werkelijkheid, die op deze vereenvoudiging van de werkelijkheid berusten, niet steeds veel succes zullen opleveren. Er is sprake van een dubbele reductie: eerst wordt de werkelijkheid sterk vereenvoudigd, getransformeerd, in getalsmatige abstracties ondergebracht, en vervolgens worden de conclusies van de onderzoekingen 'losgelaten' op de werkelijkheid, die dan wederom enigszins versimpeld moet worden om nog adequaat de onderzoekresultaten te kunnen 'inpassen' of verwerken. De teleurstelling van menig opdrachtgever bij het zien van de voornaamste conclusies van onderzoek laat zich in dit licht analyseren.

Mijn standpunt gaat uit van een logische redenering: voor een adequate toe-

passing van onderzoekresultaten is een zo nauwkeurig mogelijk beeld nodig van de werkelijkheid. Welnu, als onderzoek slechts een zeer onnauwkeurig beeld van de werkelijkheid oplevert, dan moet men zich niet verbazen over het feit dat de resultaten van dergelijk onderzoek ook onnauwkeurig aansluiten bij de 'levende' werkelijkheid.

Dit argument luistert nauw. Het is geen pleidooi tegen kwantificering of abstractie. Elk wetenschappelijk onderzoek veronderstelt vereenvoudiging en reductie. Het is een pleidooi om de reductie en de abstractie vooral te zoeken in de reikwijdte van het geobserveerde en niet in de *hoeveelheid* kenmerken van hetgeen geobserveerd wordt.

Als men gedetailleerder dan op dit moment gebruikelijk is wil gaan waarnemen, dan zal een beperking liggen in de uitgebreidheid van de onderzochte personen en niet in de uitgebreidheid van de te onderzoeken 'variabelen'. Men laat bij deze opzet uiteraard enige veren vallen. Maar veren zijn nog geen steken. Als men wéét welke veren men laat, dan is het verlies te overzien. De winst zou kunnen liggen in de betere observatie en de daarop gebaseerde verbeterde theorievorming. Het zou om die reden ook van nut kunnen zijn voor de toepassingen van onderzoek en zelfs voor het toegepaste onderzoek. Immers, niets is zo praktisch als een goede theorie. Dat is algemeen erkend, maar het feit dat goede theorievorming gebaseerd is op gedetailleerde waarneming wordt veel minder erkend.

3. EXTENSIEF EN INTENSIEF ONDERZOEK MET ELKAAR VERGELEKEN

Enkele voorbeelden van intensief onderzoek in de sociologie van de laatste tijd zijn R. Sennett and J. Cobb, *The hidden injuries of class* (1973) waar de auteurs dertien arbeidersgezinnen gedurende anderhalf jaar gevolgd hebben, in allerlei situaties hebben meegemaakt en hebben geobserveerd. Een voorbeeld uit Engeland is het boek van D. Marsden and E. Duff, *Workless* (1975) over de beleving van werkloosheid. Ook hier werd gekozen voor slechts een tiental gezinshoofden en hun nabije omgeving. Men kan zich hierbij wel afvragen of een aantal van tien respondenteenheden niet wat àl te willekeurig klein is. Recentelijk werd ook bekend het werk van de Amerikaanse sociologe Rubin met haar in het Nederlands vertaalde onderzoeken *Pijn en moeite* en *Vrouwen van zekere leeftijd*. Rubin poogt diepte-interviews te combineren met een zekere getalmatige abstractie, die zij in een aanhangsel toevoegt. Hier gaat het om 100 – 130 respondenten. Intensief onderzoek wordt uiteraard ook gevonden in de bekende participerende observatie-studies in de sociologie. In de antropologie was en is deze manier van onderzoek doen gebruikelijk. Het is zelfs zo dat het verschil tussen sociologen en antropologen wel eens is weergegeven in de houding, die beide groepen innamen tegenover hun onderzoek. Van antropologen werd gezegd dat zij zes jaar onderzoek deden (veldwerk) om dit daarna in zes maanden te publiceren. Van sociologen werd gekscherend gezegd dat zij in zes maanden hun veldwerk deden om er daarna zes jaar over te doen om het rapport te schrijven. Van marktonderzoekers kan gezegd worden dat zij in zes weken veldonderzoek doen en in *zes* dagen (moeten) rapporteren!

De hernieuwde interesse in participerende observatie in het sociologisch on-

derzoek heeft ook te maken met het veelvuldiger toepassen van antropologische methoden in onderzoek in westerse samenlevingen.

Toch zijn deze nieuwe methoden, die zich onder vele namen aandienen, niet zonder problemen. Er zijn ook vele kentheoretische verschillen tussen bijvoorbeeld etnomethodologisch, etnografisch, stadsetnografisch, symbolisch interactionistisch, fenomenologisch, kwalitatief onderzoek, zodat het onder één noemer brengen van al deze methoden onder de benaming kwalitatief onderzoek onjuist zou zijn. In het afzetten tegen het standaardtype van onderzoek kan slechts voor een heel korte periode enige gelijkenis gevonden worden.

Kentheoretische en methodologische bezinning op deze nieuwe ontwikkelingen is derhalve geboden. Welke problemen doen zich voor, wanneer Rubin zegt dat zij na het negenendertigste, zelf uitgevoerde interview bijna geen nieuwe informatie over haar onderwerp meer krijgt? Het kan zijn dat negenendertig intensieve vraaggesprekken over hetzelfde onderwerp bij de onderzoeker/ster inderdaad een beeld teweegbrengt dat enige samenhang gaat vertonen en theoretische aanknopingspunten biedt. Maar hoe geldig en hoe betrouwbaar is dit langzaam gegroeide beeld? Zijn de toetsingscriteria geldigheid en betrouwbaarheid op dit type onderzoek nog van kracht en zo ja, met welke consequenties? Is een combinatie van explorerend-intensief onderzoek met daarna een toetsend-extensief onderzoek de aangewezen richting? Maar hoe kun je een vrij gecompliceerd beeld (theorie) van de werkelijkheid toetsen door middel van een sterk gereduceerd beeld, in casu het extensieve onderzoek? Deze en soortgelijke vragen zullen in de komende tijd aan de orde moeten komen en het betekent een verlevendiging van het methodologische debat in de sociale wetenschappen.

Hier wil ik volstaan met een systematische beschrijving van de verschillen tussen beide onderzoektypen. Ik treed hierbij in het voetspoor van de wetenschapsfilosoof/sociaal-psycholoog R. Harré, die zowel de wetenschapsfilosofische grondslagen van dit type van onderzoek ontwikkelt alsmede zelf praktisch onderzoek doet op verschillende onderwerpen (o.a. op het gebied van het kopen van auto's en de voorkeur voor bepaalde automerken!)

Harré heeft het volgende overzicht gemaakt van de verschillen tussen de twee typen:

Extensief onderzoek	*Intensief onderzoek*
1. Groot aantal gevallen met enkele kenmerken.	1. Enkele gevallen met een groot aantal kenmerken.
2. Steekproefproblemen.	2. Geen steekproefproblemen, want de populatie is (nog) niet bekend.
3. Slechts waarschijnlijkheidsuitspraken.	3. Geen waarschijnlijkheidsuitspraken.
4. Meer algemeen.	4. Minder algemeen.
5. De geobserveerde klasse is scherp omschreven (gedefinieerd); de kenmerken van de klasse worden echter niet verklaard.	5. Bijzondere gevallen intensief bestudeerd en het probleem is, uit te maken van welke klasse de personen lid zijn.

6. Abstractie door middel van generalisatie.	6. Generalisatie door middel van abstractie.
Praktische verschillen	
7. Snel	7. Langzaam.
8. Rigide (gefixeerde categorieën).	8. Flexibel.
9. Laag risico: er kunnen altijd enige resultaten, zelfs als zij niet significant zijn, gerapporteerd worden.	9. Hoog risico: de totale mislukking van een conceptueel schema is mogelijk.
10. Publicatiemogelijkheden	10. Publicatie moeilijker.

(bron: Harré, 1976 en 1979).

Naast deze door Harré genoemde verschillen zou men de lijst nog kunnen aanvullen met de manier waarop het onderzoek kan worden georganiseerd: het extensieve onderzoek leent zich voor routinisering en bureaucratisering, het intensieve niet. Het extensieve onderzoek kan worden uitgevoerd door gehuurde en middelbaar geschoolde interview(st)ers, het intensieve onderzoek zal bij voorkeur en bijna altijd binnen een klein team door de onderzoek(st)ers zelf verricht worden. De voorkeur-apparatuur van het extensieve onderzoek is de computer, van het intensieve onderzoek de bandrecorder of misschien zelfs het filmopname-apparaat.

De mijns inziens belangrijkste verschillen liggen in de manier waarop gegeneraliseerd wordt en geabstraheerd van de werkelijkheid. In het extensieve onderzoek wordt geabstraheerd van concrete personen ten gunste van een plaatje, een beeld van 'de gemiddelde' persoon, de gemiddelde Nederlander, vrouw, visser, scholier, WAO-er etc. Dit is een zeer abstract beeld; waar slechts in geaggregeerde zin iets mee gedaan kan worden. De abstractie berust op *statistische generalisatie* en de geldigheid van conclusies wordt beheerst door de regels van de inductieve statistiek.

Bij intensief onderzoek wordt weliswaar óók geabstraheerd en gegeneraliseerd (de bewering dat bij 'kwalitatief' en soortgelijk onderzoek de 'levende' werkelijkheid beschreven wordt en 'levensecht' onderzoek gedaan wordt, berust op een filosofische blunder). Maar het doel van de generalisatie bij intensief onderzoek is *niet* de generalisatie *naar de andere en met name niet-onderzochte leden van de klasse*, maar juist de generalisatie *naar andere situaties, waarin de onderzochte personen zullen reageren*. Dit geschiedt door het maken van een 'mental mapping', een kaart van alle stereotiepe kenmerken en reactiewijzen van de onderzochte personen in de onderzochte situatie. Via deze 'mapping' hoopt men te komen tot het beschrijven en vervolgens verklaren van gedragingen in specifieke situaties: in die en die situatie gedragen deze mensen zich op deze manier en dat komt vanwege dit en dat. Het gaat er hierbij dus veeleer om stereotiepe gedragswijzen, vastgelegde manieren van doen, denken, willen en voelen te beschrijven en te verklaren, niet vanuit een eenvoudige statistische correlatie, maar juist vanuit een begrijpelijke redegeving van het eigen gedrag. Mensen gedragen zich niet op een bepaalde wijze ómdat ze jonger dan dertig zijn, vrouw zijn of katholiek, maar vanwege andere moeilijker via extensief onderzoek te achterhalen motieven. Alleen indien specifie-

ke gedragingen specifiek met extensieve kenmerken verbonden zijn, bijvoorbeeld de kinderen laten dopen ómdat men katholiek is, krijgen deze kenmerken verklaringswaarde. De bij het extensieve onderzoek gebruikelijke correlatie-analyse wordt daarmee niet overbodig, maar kan gezien worden als een soort hulponderzoek, waarin de richting van het eigenlijke verklarende onderzoek op intensieve wijze kan worden afgetast.

De combinatie van extensief onderzoek met intensief onderzoek wordt in mijn visie anders opgevat dan in de heersende methodologieboeken: extensief onderzoek vormt een goede voorbereiding voor het eigenlijke, verklarende – en intensieve – onderzoek en *niet andersom*.

Dat deze verklaringen op hun beurt weer getoetst zullen moeten worden op hun generalisatiemogelijkheden naar niet-onderzochte personen is een probleem dat in de komende periode serieus onderzocht zal moeten worden.

Concluderend zou ik willen zeggen dat de winst van het intensieve onderzoek in vergelijking met het extensieve onderzoek vooral gelegen is in de grotere gedetailleerdheid van de onderzochte verschijnselen (personen, situaties, redeneerpatronen, motivaties e.d.) en in de realistischer conclusies die men uit de observaties kan trekken. Dat deze conclusies niet naar alle Nederlanders te generaliseren zijn is een gegeven, dat ik voor het sociaal-wetenschappelijk onderzoek niet als een groot gemis wil zien. Leerde niet reeds Mannoury ons dat de spreiding van woordbetekenissen veel groter is dan wij geneigd zijn aan te nemen? Met andere woorden liever een grondig inzicht in een beperkt onderzochte groep dan een oppervlakkig inzicht in een statistisch tot de gehele populatie uit te breiden eigenschap van een groep.

Het is de vraag of het marktonderzoek deze stelling zou willen onderschrijven. Voor de voortgang en vooruitgang van de wetenschap is zij mijns inziens echter onontbeerlijk.

BIBLIOGRAFIE

1. G. Mannoury, 1949. *Significa, Servire,* Den Haag, p. 50.
2. R. Harré, 1963. *An introduction to the logic of the sciences,* London.
3. R. Harré, 1976. *The personality of individuals,* in: R. Harré (ed) Personality, Oxford.
4. Zie C.J.M. Schuyt, 1982. *Kennis van de sociale werkelijkheid of sociale werkelijkheid van de kennis,* in: C. Bouw, F. Bovenkerk, K. Bruin en L. Brunt (red), Hoe weet je dat, wegen van sociaal onderzoek, Wetenschappelijke Uitgeverij, Amsterdam, p. 38-52.
5. R. Harré and P.F. Secord, 1972. *The explanation of social behavior,* Oxford.
6. R. Harré, 1976. *Buying a car,* in: R. Harré (ed) Life sentences, aspects of the social role of language, Oxford.
7. Zie R. Harré, *The personality of individuals,* op cit. p. 224 en R. Harré, 1979. *Social Being,* Oxford, p. 132-137.

8. Decision-Makers onderzocht

L. ZÖLLNER

SAMENVATTING

Het belang van de Decision-Makers als afnemers van industriële, of met de bedrijfsvoering verband houdende, produkten werd al lang onderkend. Het is verheugend te zien hoe de aanzet van het FIMS onderzoek via DMS '80 uitmondt in een geoptimaliseerd onderzoeksmodel DM '82. Geoptimaliseerd niet alleen in de wijze waarop de steekproef is samengesteld, maar ook in de ontwikkeling van de samenwerkingsverbanden tussen belanghebbenden.

Een samenwerking die resulteert in een qua opzet en uitvoering beter eindresultaat. Bovendien zal DM '82 niet langer uitsluitend toegankelijk zijn voor de intiatiefnemers, doch zal het gebruik van dit onderzoek ook open staan voor andere uitgevers van relevante media.

Een grotere toepasbaarheid en toenemende kwaliteit zullen ongetwijfeld leiden tot een toenemende autoriteit van de Decision-Makers Onderzoeken in Nederland. En dat beantwoordt volledig aan de lange-termijndoelstelling van de initiatiefnemers.

1. ONDERZOEK IN ONTWIKKELING

Mag het soms zo lijken dat onderzoek in Nederland en daarbuiten veelal verloopt langs de welbekende en platgetreden paden, er zijn wel degelijk ook gebieden waarin werkelijk innovatief te werk wordt gegaan.

Eén van die gebieden vormt het onderzoek onder beslissingsbevoegde functionarissen in het bedrijfsleven en bij de overheid.

In deze bijdrage worden de doelstellingen en methoden van onderzoek op dit terrein in een historisch perspectief geplaatst. Een perspectief waaruit een duidelijke kwaliteitsverbeterende ontwikkeling valt af te leiden.

2. DECISION-MAKERS

Onze Britse buren kunnen het vaak zo kernachtig uitdrukken. Dat geldt ook voor de hier inmiddels ingeburgerde term 'Decision-Makers' als dynamisch klinkend equivalent voor 'beslissingsbevoegde functionarissen'.

Helemaal een equivalent is het echter ook weer niet. Ligt bij de Nederlandse term de nadruk op de *bevoegdheid* tot het nemen van beslissingen, het Engelse begrip Decision-Maker gaat voorbij aan dit beperkende criterium.

Tot Decision-Maker wordt iedere persoon gerekend die bij besluitvormingsprocessen betrokken is.

En in die betekenis wordt het begrip ook in de Nederlandse situatie gehanteerd.

Decision-Makers zijn dus personen die vanuit hun functioneren in bedrijf of instelling, betrokken zijn bij beslissingsprocessen die voor de bedrijfsvoering van belang zijn.

Deze omschrijving is er één uit vele. In de wetenschappelijke literatuur rond

het begrip 'Decision-Making' doen er diverse de ronde, alle op min of meer subtiele wijze verschillend van elkaar.

De hier gebruikte omschrijving is ruim genoeg om ze alle te omvatten en is daarom bruikbaar voor de begripsvorming. Zoals hierna zal blijken, echter niet adequaat voor onderzoek bij deze groep.

Het fenomeen Decision-Maker geniet het laatste decennium warme belangstelling van uiteenlopende disciplines. Zo zijn sociale wetenschappers en psychologen geïnteresseerd in de intermenselijke relaties en afhankelijkheidspatronen, leidersrollen en beïnvloedingsprocessen in een sociaal systeem als het arbeidsveld.

Dan zijn er de opleidingsinstituten die zich toeleggen op de training van toekomstige middelbare en hogere functionarissen in het bedrijfsleven, die Decision Making (terecht) als een essentieel onderdeel van de opleiding zien.

Ten slotte is er de omvangrijke en weinig gestructureerde groep aanbieders van produkten en diensten voor het bedrijfsleven. Zelf vaak Decision-Makers, hebben zij er uit commercieel oogpunt belang bij inzicht te krijgen in de totstandkoming van aankoop- en investeringsbeslissingen. Hoewel de interesse groot was en het belang van een gedegen inzicht onderkend werd, konden deze producenten/dienstverleners kennelijk nooit de handen op elkaar krijgen voor enig houtsnijdend onderzoek op dit gebied.

De belangen lagen daarvoor wellicht te ver uit elkaar, of juist te veel op dezelfde lijn. De directe revenuen waren wellicht niet voor ieder meteen zichtbaar, óf de belangstelling in de besluitvormingsprocessen beperkte zich te veel tot één bepaalde groep afnemers, waarvan aan aanbiederszijde al het een en ander bekend was over de wijze waarop aankoopbeslissingen tot stand komen. Hoe het ook zij, een vierde groep belanghebbenden diende zich aan. De uitgevers van persmedia. Met name de uitgevers van landelijke dagbladen en van tijdschriften waarvan de advertentiebezetting voor een belangrijk deel bestaat uit 'industriële' adverteerders, toonden een open oog voor de waarde van een helder inzicht in de besluitvormingsprocessen in het bedrijfsleven. Zij onderkenden de bestaande onzekerheid aan adverteerderszijde ten aanzien van de kwaliteit van het ruime aanbod van persmedia op managementgebied. Media die elk voor zich het meest effectieve bereik in de industriële doelgroep pretendeerden te bieden.

Daarnaast vormden voor de uitgevers de in aantal en omvang toenemende stroom 'below the line'-activiteiten en de in het buitenland gesignaleerde groei van de industriële markt als zodanig, de aanleiding om adverteerders tegemoet te willen treden met helder en vergelijkbaar onderzoekmateriaal omtrent de aard en omvang van de groep Decision-Makers. Dat in dit type onderzoek het lezen van media uitgebreid aan de orde komt zal niemand verbazen. Het was de uitgevers er uit commercieel oogpunt om te doen duidelijk te maken welke media zich in een hoge populariteit van de Decision-Makers mogen verheugen. De eerste poging in deze richting werd ondernomen door Dagbladuitgever De Telegraaf, enige tijd later gevolgd door Elsevier-NDU's werkmaatschappij Bonaventura, uitgever van o.a. Elseviers Magazine en FEM, Financieel Economisch Magazine. Een derde Decision-Makers-onderzoek is begin september van dit jaar verschenen op initiatief van beide uitgeversmaatschap-

pijen. Het is interessant de evident aanwezige ontwikkelingslijn van deze drie onderzoeken te volgen. Doelstelling, methode van steekproeftrekking en definiëringsverschillen zullen hierbij specifieke aandacht krijgen.

3. DECISION-MAKING UNIT

De wens om Decision-Makers te onderzoeken wordt direct gevolgd door de confrontatie met het probleem om vast te stellen wie hiertoe gerekend mogen worden. De literatuur op dit gebied wijst terecht op het feit dat er, zeker in de huidige bedrijfsorganisaties, niet zoiets bestaat als één Decision-Maker, doch dat de besluitvormingsprocessen juist tot stand komen door de inbreng van meerdere personen. Een *groep* functionarissen die niet alleen in hiërarchische zin, maar ook in de mate van betrokkenheid grote diversiteit vertoont, maakt als collectief een beslissingsprocedure door die resulteert in een aankoop- of investeringsbesluit. De moeilijkheid op onderzoektechnisch gebied ligt hierin dat dit beslissingscollectief geen statisch geheel is, doch zich kenmerkt door een grote verscheidenheid en dynamiek op tal van aspecten.

Zo kan het aantal personen waaruit een Decision-Making Unit is samengesteld variëren als gevolg van onder meer:

- de omvang van het bedrijf, gemeten in het aantal personeelsleden;
- de organisatorische structuur van de onderneming (relatief veel/weinig inspraakprocedures);
- de aard van de onderneming (veel/weinig vakspecialisten in huis);
- de aard van het produkt (aanschaf kantoorpand versus inkoop nieuwe schrijfmachine).

Het laatste voorbeeld geeft aan dat de DMU ook qua functieniveau een grote verscheidenheid kan vertonen.

En ten slotte is er het verschil in beslissingsinvloed van personen binnen de DMU. Een secretaresse zal als lid van een gevormde DMU weliswaar gekend worden bij de aankoopbeslissing van een nieuwe schrijfmachine, het *gewicht* van haar stem in de DMU zal echter een andere zijn dan die van het hoofd inkoop.

Onduidelijker nog zijn de officieuze, niet in organisatieschema's vast te leggen beïnvloedingstrajecten binnen een onderneming. Te denken is hierbij aan vertrouwensrelaties, externe adviseurs, bestaande verbintenissen met toeleveranciers etc.

Geconfronteerd met dit amoebe-achtige verschijnsel van de DMU, werden de aanvankelijke plannen tot het in kaart brengen hiervan losgelaten en is in de inmiddels uitgevoerde onderzoeken steeds uitgegaan van de afzonderlijke personen die bij beslissingsprocessen betrokken zijn, de Decision-Makers.

Maar ook dan volgt een probleem, want wie dient als Decision-Maker aangemerkt te worden? Enig nadenken leidt tot de conclusie dat vrijwel iedereen die in het arbeidsproces is opgenomen op enigerlei wijze wel eens bij een aankoopbeslissing betrokken is.

Vasthouden aan deze definiëring van een Decision-Maker zou betekenen dat doelgroepgericht onderzoek overbodig zou zijn en dat men zich beter tot een algemeen bevolkingsonderzoek als het Nationaal Onderzoek Persmedia

kan wenden. De praktijk biedt echter uitkomst. Daarin blijkt namelijk dat er binnen elke organisatie wel degelijk personen aan te wijzen zijn aan wie meer dan aan anderen een stem wordt toegedacht bij het nemen van beslissingen.

Het is dus vooral een kwestie van het stellen van een ondergrens, op grond waarvan iemand al dan niet als Decision-Maker aangemerkt kan worden. Het is met name de bepaling van die ondergrens die een vergelijking van de drie verschenen onderzoeken op dit gebied interessant maakt.

Het eerst verschenen onderzoek: FIMS: Financieel Industrieel Marketing Service, gaat uit van een brede samenstelling van de DMU. En richt zich als gevolg daarvan op een zeer brede groep werknemers.

Het DECISION-MAKER SURVEY dat in 1980 verscheen, richt zich als gevolg van de strenge selectiecriteria uitsluitend op de kleine groep topmanagers/ Decision-Makers. Het laatst verschenen onderzoek, DM '82, vormt een synthese van deze twee en is ook qua selectieprocedure volledig afwijkend van de voorgaande.

4. FIMS 1977

4.1. *Doelstelling*

Dagbladuitgeverij De Telegraaf komt de eer toe als eerste een poging te hebben ondernomen om de groep Decision-Makers in Nederland in kaart te brengen. In september 1977 presenteerde zij het FIMS-onderzoek, Financieel Industrieel Marketing Service.

Aanleiding tot dit onderzoek vormde het gesignaleerde gebrek aan meer dan slechts oppervlakkige informatie uit bestaand (NOP)onderzoek, over doelgroepen die van belang zijn voor de financieel-economische en industriële markt.

De primaire doelstelling van het onderzoek was dan ook meer inzicht te geven in het socio-economische profiel, de verantwoordelijkheden en het leesgedrag van personen waarop de financieel-economische berichtgeving zich richt. Met name dat van De Financiële Telegraaf. Daarnaast echter is ook de betrokkenheid van personen gemeten bij de aanschaf van goederen en diensten, de Decision-Makers dus.

4.2. *Steekproef FIMS*

Vooronderzoek leerde dat ca. 90% van de personen waarop financieel-economische berichtgeving zich richt, behoort tot de A/B1/B2 welstandsklassen. Het onderzoekbureau Research en Marketing B.V. te Heerlen beschikte over een basissteekproef van 600 A/B adressen. Twee adressen ter linker- en rechterzijde van het startadres fungeerden, indien deze aan de A/B criteria voldeden, als interviewadres. In totaal werden 3200 vraaggesprekken gevoerd, hetgeen uiteindelijk resulteerde in een nettosteekproef van 2.517 personen. Ondervraagd werden gezinshoofden vanaf 25 jaar. De in het onderzoek gevonden resultaten werden vervolgens geprojecteerd op het Nederlandse universum. Volgens NOP '75 omvatte de groep gezinshoofden 25+ in de welstands-

klasse A/B 1.468.000 personen, hetgeen ca. 29% vertegenwoordigde van de toenmalige totale mannelijke bevolking in Nederland. Er vond, voorafgaande aan het eigenlijke interview, *geen* voorselectie plaats middels vragen waarmee de eventuele betrokkenheid bij beslissingsprocessen kon worden vastgesteld.

Het hanteren van de welstandsklassen A, B1 en B2 als basissteekproef zonder verdere voorselectie, brengt met zich mee dat FIMS primair inzicht geeft in het leesgedrag van personen in de A/B klassen en pas in de tweede plaats in het leesgedrag van Decision-Makers.

Onder Decision-Makers verstaan we binnen het kader van FIMS dan, alle personen die opgaven een adviserende, medebeslissende of formeel bepalende invloed te hebben bij de aanschaf van ten ministe één uit een serie produkten/ diensten.

Deze werkwijze leidde tot tamelijk verbluffende uitkomsten. Zo zou 52% van de steekproef, ofwel 774.000 personen, betrokken zijn bij de selectie of aanstelling van personeel. Bijna 30%, ofwel 443.000 personen, bij de aanschaf van bedrijfspanden. En ng eens 28% (410.000) bij het afsluiten van verzekeringen. Op deze wijze geformuleerd en geprojecteerd op het Nederlands universum lijkt de groep Decison-Makers uit een ongedacht groot aantal te bestaan.

Het valt niet moeilijk te zien dat dit het rechtstreekse gevolg is van de ruim gedefinieerde selectiecriteria, waarbij geen voorselectie is gemaakt en waarbij op generlei wijze onderscheid is gemaakt naar de *mate* van betrokkenheid tijdens de verschillende fasen van het aankoopproces (ideevorming, oriëntatie, beslissingsfase). Ook wordt voorbijgegaan aan het verschil in gewicht van personen binnen de DMU.

De omvang van de groep kan theoretisch gezien wel juist zijn, in de praktijk blijkt er een duidelijk verschil te bestaan in beïnvloedingscapaciteit of beslissings'power' tussen de leden van een DMU. Het achterwege laten van een nadere voorselectie heeft als belangrijkste voordeel dat hiermee een grote uitval van personen in de basissteekproef voorkomen wordt. Het geeft daardoor op zichzelf een goed beeld van de *omvang* van de groep 'potentiële' Decision-Makers. Een brede steekproefopzet verkleint de kans op verlies van een in aantallen personen gemeten belangrijk deel van de DMU, doch levert als resultaat een enigszins verwaterd beeld op.

Criteria zó ruim gedefinieerd hebben nauwelijks nog discriminerende waarde. Indien ook de welstandsklasse C in het onderzoek betrokken zou zijn, zou ook een fors deel van deze grote groep werknemers zonder problemen in de steekproef kunnen worden toegelaten.

Hier stuiten we op het al eerder genoemde dilemma dat kenmerkend is voor dit type onderzoek. Waar leggen we de grens bij de bepaling of iemand al dan niet Decision-Maker is? Het is een dilemma dat slechts overwonnen kan worden op grond van praktische overwegingen ten aanzien van de bruikbaarheid van de onderzoekresultaten.

Daartegenover staat het voorbeeld van strenge selectiecriteria, zoals gehanteerd in het DMS-onderzoek van Bonaventura. Een selectieprocedure waarbij bewust gekozen is voor een verlies in de steekproef van personen die slechts incidenteel betrokken zijn bij beslissingsprocessen.

Afgezien van deze kanttekeningen verdient het FIMS-onderzoek zeker

waardering, omdat het de eerste serieuze poging was tot het verkrijgen van meer inzicht in de aard en omvang van een belangwekkende doelgroep, waaronder Decision-Makers. FIMS is als zodanig ook bijzonder waardevol geweest, met name voor adverteerders en reclamebureaus. Deze hoefden zich niet langer te verlaten op optimistische verhalen van media-exploitanten, doch konden op onderling vergelijkbare wijze werken met bereikscijfers van media, hetgeen de kwaliteit van de mediaplanning op dit gebied aanzienlijk heeft verbeterd.

5. DMS 1980

5.1. *Doelstelling*

In het voorjaar van 1980 ondernam de uitgever van Elseviers Magazine, de B.V. Uitgeversmaatschappij Bonaventura, een tweede poging tot een inventariserend bereiksonderzoek onder Decision-Makers.

In tegenstelling tot het FIMS-onderzoek lag de primaire doelstelling van Bonaventura niet in de analyse van een zo groot mogelijk aantal personen binnen de DMU, doch in de verkenning van dát gedeelte binnen de DMU dat ook in kwalitatieve zin tot de Decision-Makers van het meeste gewicht gerekend mag worden.

Het Decision-Maker Survey 1980 is dan ook een onderzoek dat betrekking heeft op de werkelijk veelvuldig bij beslissingsprocessen betrokken functionarissen, de TOP DECISION-MAKERS.

Dat deze categorie aanzienlijk kleiner zou uitvallen dan in het FIMS werd aangetroffen lag hierbij in de lijn der verwachtingen. Welke omvang de groep van Top Decision-Makers in Nederland dan wél zou hebben bleef tot op het tijdstip van afronding van het veldwerk een open vraag.

5.2. *DMS methode van onderzoek*

Bij de bestudering van methoden om de gezochte doelgroep op te kunnen sporen, waarbij binnen- en buitenlands onderzoek tot voorbeeld diende, bleek alras dat de gevolgde technieken slechts twee soorten resultaten opleverden. Ruime definiëringen werkten als een zeef met te ruime mazen, met een zich niet bijster onderscheidende steekproef als resultaat. Strakke definities, bijvoorbeeld alleen welstandklasse A onderzoeken, hadden steevast tot gevolg dat ook nogal wat personen in de steekproef belandden die vanwege hun functie niet of nauwelijks bij beslissingsprocessen betrokken zijn. Uiteindelijk werd daarom gekozen voor een *samenstelling* van selectiecriteria, met als doel te komen tot de formatie van een steekproef die in overgrote meerderheid zou bestaan uit de beoogde groep Decision-Makers. Hierbij werden de volgende uitgangspunten gehanteerd. De bedoelde groep zou vrijwel uitsluitend te vinden zijn in de categorie personen die volgens de WBO-welstandsindeling behoren tot de klassen A/B1. Deze groep omvat de hoogste 25% van de Nederlandse bevolking. (Toekenning van een welstandsklasse geschiedt op objectieve gronden op basis van opleiding, beroepsgroep en in mindere mate ook leeftijd).
Uitgangspunt 2.

Er werd van uitgegaan dat het beroep en de functie wezenlijke indicatoren vormen voor de vaststelling of de persoon in kwestie al dan niet tot de gezochte toplaag van Decision-Makers zou behoren. Binnen de welstandsklassen A/B1 bevinden zich nogal wat beoefenaren van beroepen die overduidelijk *geen* beslissingsbetrokkenheid kennen als bedoeld in het kader van dit onderzoek. Bijvoorbeeld: leraren, dominees, huisartsen en studenten.
Uitganspunt 3.

De op basis van welstandsklassen en beroepsselectie geformeerde steekproef zou een controle moeten ondergaan met betrekking tot het daadwerkelijk en ook als zodanig ervaren van 'betrokkenheid' bij aankoop- en investeringsbeslissingen. Een controle dus waarbij de incidenteel betrokkenen zoveel mogelijk zouden worden uitgezeefd.

5.3. *Steekproeftrekking*

Fase 1.

De basis steekproef van A/B1 adressen werd samengesteld uit 8 weekenquêtes van de NSS-omnibus (waaruit 2.297 adressen werden geselecteerd) en uit 3 kwartaalenquêtes die de N.V. v/h Nederlandse Stichting voor Statistiek te 's Gravenhage in 1979 liet uitvoeren. Hieruit werden 2.144 adressen geselecteerd. Deze twee doelsteekproeven werden op vergelijkbaarheid gecontroleerd en toonden geen statistisch belangrijke afwijkingen.

Fase 2.

De basissteekproef van 4.413 A/B1 adressen werd vervolgens gescreend op basis van het beroep of de functie van de hoofdkostwinner. In totaal werden zo 1.653 potentiële Decision-Makers aangetroffen. Bij twijfel omtrent het beroep in relatie tot beslissingsbetrokkenheid werd het adres in de steekproef opgenomen. Op voorhand werden uitgesloten: gepensioneerden zonder beroepsaanduiding, artsen (behalve geneesheer-directeur), leraren, ambtenaren met een rang lager dan referendaris, militairen met een rang lager dan kolonel.

Ter controle op deze procedure werd uit de groep van uitgeslotenen een controlesteekproef samengesteld van 220 adressen. Geen van deze personen bleek betrokken te zijn bij investeringsbeslissingen, terwijl slechts 10% aangaf wel eens betrokken te zijn bij de aankoop van goederen. Dit resultaat gaf voldoende zekerheid dat de selectie op beroep geen al te rigoreuze procedure was.

Fase 3. Controlevragen.

De 1.653 geselecteerde adressen werden door enquêteurs van de NSS bezocht. Voordat het eigenlijke interview een aanvang nam diende de ondervraagde persoon eerst positief te antwoorden op minstens één van de volgende twee controlevragen.

A. Houdt u zich op grond van uw functie of positie bezig met het investeren of beleggen van geldmiddelen van een bedrijf of instelling, bijvoorbeeld als adviseur, stafmedewerker of directielid e.d.?
B. Houdt u zich op grond van uw functie of positie bezig met de keuze en aanschaf van kapitaalgoederen voor een bedrijf of instelling, bijvoorbeeld als adviseur, stafmedewerker, directielid e.d.?

5.4. *Pilot*

In de pilotfase van dit onderzoek werd aan deze twee selectievragen nog een derde toegevoegd, t.w. de vraag of men geïnteresseerd was in informatie over bedrijven van instellingen.

Evenals bij het FIMS-onderzoek bleek deze vraag een negatief effect op de selectiviteit te hebben. Als een soort spons zuigt dit type vragen personen op in de steekproef die er op grond van alle andere kenmerken helemaal niet thuis horen.

5.5. *Netto-steekproef*

Dat de gehanteerde twee selectievragen hun werk deden mag blijken uit het feit dat van de bereikte *potentiële* Decision-Makers (excl. non-respons) bijna de helft zichzelf niet herkende in één van beide selectievragen.

Het valt inderdaad niet te ontkennen dat de toonzetting en letterlijke betekenis van deze vragen alleen hen aanspreekt die inderdaad hoog-frequent bemoeienis hebben met aankoop- en investeringsbeslissingen.

De ontwikkeling van bruto- naar netto-steekproef verliep als volgt:

NSS omnibus A/B1 adressen	2.297
NSS kwartaalenquête A/B1 adressen	2.144
Totaal	4.441
Selectie op beroep/functie (potentiële Decision-Maker)	1.653 = 100%
Na bezoek: Decision-Makers	576 = 35%
geen Decision-Makers	537 = 33%
non-respons	348 = 21%
onbruikbare adressen	192 = 12%

Slechts ruim een derde van de op beroep voorgeselecteerde personen bleek 'bestand' tegen de twee controlevragen. Een te strenge selectie? Niet als men de doelstelling van dit DMS onderzoek in beschouwing neemt. Bewust is gekozen voor een onderzoek onder Top Decision-Makers. En die waren langs deze weg wel degelijk bereikt, getuige de onderzoekuitkomsten.

45% van hen bleek als zelfstandige werkzaam te zijn.

38% van hen die in loondienst waren, bleek een functie als directeur te vervullen.

Hetgeen aangeeft dat Bonaventura met dit onderzoek inderdaad het neusje van de Decision-Makers-zalm te pakken had.

Een niet voorziene consequentie van de gevolgde steekproefprocedure trad aan het licht toen de eerste resultaten bekend werden. Ruim 1/5 van de steekproef bleek te bestaan uit personen die werkzaam zijn in de agrarische sector! Hoewel het geenszins in de bedoeling lag deze sector geheel uit te sluiten, was dit resultaat niet geheel in overeenstemming met hetgeen Bonaventura verwachtte ten aanzien van de verdeling van de doelgroep over de verschillende bedrijfstypen. Temeer niet daar landelijk gezien slechts 5% van de beroepsbevolking in deze sector werkzaam is. Zij was echter het logische gevolg van de

definitie voor de welstandsklasse A, waartoe *iedere* zelfstandige ondernemer automatisch gerekend wordt.

Afgezien van deze omissie leverde DMS 1980 een schat aan informatie op over de Top Decision-Makers in Nederland. Op grond van berekeningen van de NSS kan deze groep op een totaal van 215.000 personen geschat worden. Een aanzienlijk kleiner aantal dan via de FIMS-methode kon worden berekend.

5.6. *Resultaten*

Het DMS geeft behalve over het leesgedrag van de Decision-Maker m.b.t. een groot aantal binnen- en buitenlandse media, ook inzicht in zijn verantwoordelijkheden, zijn mate van betrokkenheid bij de aanschaf van goederen en diensten, zijn mate van interesse in redactionele onderwerpen, zijn bezitspatroon en zijn gebruikskenmerken en de gebruikelijke socio-economische criteria zoals opleidingsniveau, burgerlijke staat, regio en leeftijd. Nieuw, en de bruikbaarheid belangrijk beïnvloedend, was de conversie van het databestand naar MMTS-programmatuur. Dit computerprogrammapakket wordt in Nederland algemeen gebruikt voor mediaplanningsdoeleinden op basis van de NOP-onderzoeken. Het DMS-bestand, geschikt gemaakt voor het gebruik op MMTS, had tot gevolg dat mediaplannend Nederland voor het eerst in staat was gedegen evaluaties van mediaschema's te maken, waarmee een belangrijke kwaliteitsverbetering optrad op het gebied van mediaplanning voor industriële adverteerders. De komst van DMS had tevens tot gevolg dat een aantal uitgevers van tot op dat moment slechts vagelijk geprofileerde tijdschriften op het gebied van management, zich opeens voorzien zagen van een grote hoeveelheid informatie over hun media, hetgeen hen in het algemeen geen windeieren heeft gelegd.

5.7. *De beperkingen*

Het is natuurlijk niet alles even schoon in deze wereld. Ook niet bij het DMS. De belangrijkste beperking van het onderzoek schuilt in de omvang van de netto-steekproef. 576 personen is geen aantal waarmee nog erg veel uitsplitsingen te maken zijn zonder direct in niet-representatieve cellen te belanden. Een gegeven waaraan terecht tamelijk zwaar getild wordt. Zonder aan dit bezwaar te willen afdingen dient toch opgemerkt te worden dat het werken met (te) kleine cellen niet typisch is voor DMS. Regelmatig komt het voor dat adverteerders genoegen nemen met mediavoorstellen die gebaseerd zijn op NOP, waarbij de doelgroep waarop de mediakeuze werd vastgesteld, nog vele malen geringer in omvang is. Bijvoorbeeld: directeuren met tien man en meer personeel, waarvan er ook in NOP '79 niet meer dan 70 voorhanden zijn.

Een tweede beperking van DMS is dat het op wel zeer rigoreuze wijze voorbij gaat aan de in omvang grote groep Decision-Makers die te vinden is in het echelon direct onder dat van de top managers. Functionarissen die wellicht niet dagelijks investeringsbeslissingen nemen, doch als vakspecialisten of hoofden van dienst wel degelijk hun inbreng hebben in het besluitvormingsproces.

FIMS te breed, DMS wellicht te smal. Het werd tijd voor een synthese. En die kwam er in de vorm van het derde onderzoek op dit gebied, Decision-Makers '82.

6. DECISION-MAKERS '82

6.1. *Doelstelling*

DM '82 is tot stand gekomen in een samenwerkingsverband tussen de initiatiefnemers van beide voorgaande onderzoeken, Dagbladuitgeverij De Telegraaf en Bonaventura. Gesteund door de wetenschap van het commerciële belang van up-to-date feitenmateriaal omtrent Decision-Makers en met de ervaring opgedaan tijdens de voorafgaande projecten, moest het mogelijk zijn tot een aanzienlijke optimalisering van Decision-Makers Onderzoek te komen. Dit motief stond dan ook voorop bij de opzet van het nieuwe onderzoek. De doelstelling bleef vrijwel ongewijzigd. Adverteerders en reclamebureaus inzicht verschaffen in het leesgedrag, de verantwoordelijkheden, betrokkenheid bij aankoop- en investeringsbeslissingen en het socio-economisch profiel van de Decision-Makers in Nederland. De methode van onderzoek ondervond belangrijke wijzigingen.

6.2. *Welstandsklasse exit*

Aangezien welstandsklasse een geconstrueerd criterium is, gebaseerd op opleiding en in tweede instantie beroepsgroep, levert de samenstelling van de basissteekproef op grond van welstandsklassen altijd een aantal personen op die qua opleiding wél, doch qua beroepsgroep niét tot de DMS-steekproef behoren. Die moeten er dus in tweede instantie altijd weer uitgezeefd worden. Bovendien brengt de selectie van een beperkt aantal welstandsklassen altijd met zich mee dat een aantal beroepsgroepen buiten beschouwing worden gelaten waarin zich wel degelijk belangrijke Decision-Makers kunnen bevinden.

Onderstaande crossing geeft een goede indruk van de spreiding van welstandsklassen over de beroepsgroepen.

NOP '79
Beroepsgroepen: 1. Directeuren 10+
2. Directeuren 9−
3. Vrije beroepen
5. Hogere employees
6. Middelbare employees
(Beroepsgroep 4, boeren en tuinders, is hier buiten beschouwing gelaten.)

Mannen 20-64 jaar N = 3.274 Projectie 3.858.000

Beroepsgroepen	*1*	*2*	*3*	*5*	*6*
N =	70	168	36	212	792
Aantal × 1.000	76	187	38	233	907

Welstandsklassen					
A	89%	6%	53%	53%	5%
B1	9	23	21	29	40
B2	1	39	15	12	40
C	1	30 } 72%	11 } 26%	2 } 14%	16
D	–	3	–	–	1

Een selectie op bijvoorbeeld welstandsklassen A/B1 brengt met zich mee dat 72% van de directeuren 9– buiten beschouwing blijft, evenals 26% van de categorie vrije beroepen en 14% van de hogere employees. Een zuiverder methode leek het daarom om direct een selectie te maken van de basissteekproef op grond van *beroepsgroepen*.

6.3. *Selectievragen*

Géén selectievragen zoals bij het FIMS-onderzoek, of strenge selectievragen zoals bij DMS '80, bleek enerzijds drempelverlagend, anderzijds drempelverhogend uit te werken. Uitvoerige discussies omtrent dit probleem leidden ten slotte tot de conclusie dat het eigenlijk de voorkeur verdient de respondent zichzelf een gewicht toe te laten kennen ten aanzien van zijn belang als Decision-Maker. Dit doet hij dan door aan te geven hoe vaak en in welke mate hij betrokken is bij aankoop- of investeringsbeslissingen. Het wordt de respondent dus niet meer gevraagd of hij zichzelf als betrokkene ziet, dat blijkt achteraf, uit de mate waarin hij zelf heeft aangegeven betrokken te zijn bij de (eventueel toekomstige) aanschaf van produkten en diensten.

6.4. *Methode van onderzoek*

Uitgangspunten:
- Er bestaat een positieve correlatie tussen de functie of het beroep van de personen en de mate waarin zij bij beslissingsprocessen betrokken zijn.
- Zowel FIMS als DMS wijzen uit dat verreweg het grootste deel van de groep Decision-Makers tot het mannelijke geslacht behoort. Kostenoverwegingen te zamen met de veronderstelling dat de weinige vrouwelijke Decision-Makers, qua leesgedrag en socio-economisch profiel, grote overeenkomst zullen vertonen met hun mannelijke collega's, deden ons besluiten alleen mannen in de steekproef op te nemen.
- De steekproef moest voldoende groot zijn om ook zinvolle segmentaties op het bestand te kunnen uitvoeren. Gekozen is daarom voor een nettosteekproef van 2.000.
- Decision-Makers van enige importantie zullen zich naar alle waarschijnlijkheid bevinden in de volgende beroepsgroepen:
  1. directeuren 10+, directeuren met meer dan 10 man personeel;
  2. directeuren 9–, directeuren met 9 man of minder personeel. Teneinde de zeer brede groep van kleine detaillisten uit te zuiveren werd een ondergrens ingesteld van ten minste 3 man personeel;
  3. beoefenaren van hogere vrije beroepen;
  5. hogere employees en directeuren in overheidsdienst;

6. een selectie uit de groep middelbare employees.

Deze laatste categorie omvat een groep van bijna 900.000 werkzame mannen (NOP '79). Een analyse van de beroepen die tot deze categorie gerekend worden, leert dat zich hieronder een groot aantal bevindt waarvan met zekerheid gezegd kan worden dat zij geen beslissingsbetrokkenheid hebben. Zoals daar zijn: acteurs, apothekers-assistenten, verplegend personeel en talrijke lagere overheidsfunctionarissen. Voorafgaande aan het veldwerk is door het Onderzoekbureau R+M een selectie gemaakt van de beroepen binnen deze categorie die wél in aanmerking komen om in de steekproef te worden opgenomen.

6.5. *Methode van steekproeftrekking*

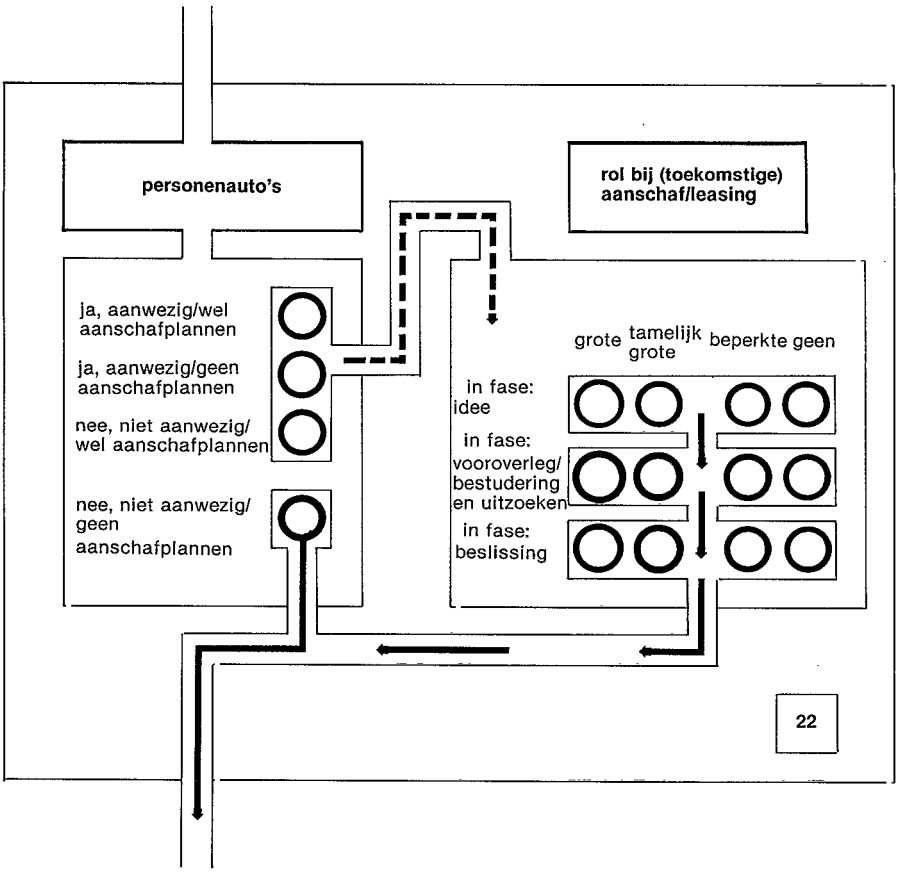
Op basis van een random getrokken steekproef van adressen werden door R+M te Heerlen op landelijke schaal telefonisch contacten gelegd met ca. 25.000 huishoudingen. Tijdens dit gesprek werden vragen gesteld omtrent het beroep, de functie en eventueel de rang van de hoofdkostwinner. De gesprekken werden in de regel overdag met de huisvrouw gevoerd. De non-respons bleef beperkt tot 22%. Indien de hoofdkostwinner aan de gestelde beroepscriteria voldeed, werd hem persoonlijk enige tijd later een interview afgenomen. Het omvangrijke en kostbare traject van de telefonische screening was nodig om een landelijk representatieve bruto-steekproef op te bouwen van ruim 2.700 respondenten.

6.6. *Het meten van beslissingsbetrokkenheid*

Zoals gezegd is bij dit onderzoek geen selectievraag toegepast met betrekking tot de classificatie als Decision-Maker. Deze classificatie vond tijdens het interview plaats op een wijze die in Nederland tot nu toe niet eerder gebruikt werd. Geïnspireerd door de in Berlijn 1980 gehouden voordracht van D. Müller-Veeh* heeft R+M een stroomdiagram ontwikkeld waarmee het de respondent mogelijk wordt gemaakt eigenhandig aan te geven of, en zo ja in welke mate, hij betrokken is of zal zijn bij de aanschaf van produkten en diensten. Het diagram waarvan hier een afbeelding is weergegeven, laat de respondent eerst per produkt of dienst afzonderlijk aangeven of het aanwezig is dan wel of daarvoor aanschafplannen bestaan. Indien dit het geval is dient hij op de rechterzijde van het diagram aan te geven in welke mate hij gedurende de verschillende fasen van het aankoopproces betrokken is of zal zijn. Daarbij worden drie fasen onderscheiden: de fase van de ideevorming, van het overleg en van de besluitvorming.

De respondent vulde eigenhandig 32 van deze stroomdiagrammen in. Voor elke produktgroep of -dienst één. Invulling vond plaats tijdens het interview na mondelinge instructie van de enquêteur.
Deze werkwijze is middels een pilot gecontroleerd.

* D. Müller-Veeh, April 1980. *Development of a self-completion media questionaire for magazines.* Berlin.



7. NABESCHOUWING DM '82

Deze methode heeft een aantal in het oog springende voordelen. De zelfwerkzaamheid van de respondent blijkt in de praktijk een positieve invloed te hebben op de kwaliteit van de gegeven antwoorden. De vermoeidheidsverschijnselen die optreden bij het mondeling beantwoorden van een lange reeks produktvragen worden hiermee omzeild. Het bleek zelfs mogelijk een aanzienlijke uitbreiding van de lijst produkten en diensten door te voeren (totaal 32 produkten en diensten). Een tweede minstens zo belangrijk voordeel is dat de respondent in staat gesteld wordt een genuanceerd beeld van zichzelf te geven met betrekking tot de betrokkenheid in de verschillende fasen van het aankoopproces. In tegenstelling tot voorgaande onderzoeken, waarbij de Decision-Maker hetzij als adviseur, hetzij als medebeslisser of als formele beslisser werd ingedeeld, is hij bij deze techniek één en dezelfde persoon.

Bovendien moet niet uitgesloten worden geacht dat met deze ondervragingsmethodiek een beter inzicht ontwikkeld kan worden in de samenstelling van de Decision-Making Unit en in het gewicht van elk der leden hiervan bij aankoopbeslissingen. Aangezien deze gegevens gekoppeld zijn aan kenmerken

als beroepsgroep, functie, verantwoordelijkheid en opleidingsniveau, kan dit nog heel wat interessante informatie opleveren. Niet alleen voor mediaplanners of adverteerders, maar ook voor sociologen en bedrijfsorganisatiespecialisten.

Het Decision-Makers Onderzoek is, voorafgegaan door een pilotstudy, in april van dit jaar het veld ingegaan. De resultaten zijn tijdens de opstelling van dit schrijven nog niet beschikbaar. Tegen de tijd dat u dit handboek in handen krijgt zal DM '82 echter gepresenteerd zijn. Dan zal ook een presentatierapport beschikbaar zijn. Het wordt u op aanvraag gaarne toegezonden.

9. Aankoopstructuren nader bezien – een onderzoek bij een drietal frequent gekochte produkten

J.P. WIJNBERGER/J. VERHULP

1. INLEIDING

Dit artikel beoogt de hoofdlijnen van een empirisch onderzoek aangaande de problematiek van het aankoopgedrag van gezinshuishoudingen weer te geven. Kernvraag bij dit onderzoek is of het mogelijk is om door middel van een analyse die aangrijpt bij de kleinste bouwstenen van het gedrag der gezinshuishoudingen, nl. 'individuele' kwantiteiten in hun volgtijdelijk verloop, een verrijkt inzicht te verkrijgen in voor het (commercieel) beleid relevante grootheden. Voorbeelden van deze laatste zijn: de in een bepaalde periode gemiddeld per gezin gekochte hoeveelheden en de daarmee samenhangende totale afzet, de gemiddeld betaalde prijzen en daarmee samenhangende totale omzet, etc.

Teneinde deze kernvraag te beantwoorden zijn voor een drietal frequent gekochte huishoudelijke produkten, te weten fijnwasmiddelen (FW, standaardhoeveelheden van 920 gr), cola (C, standaardhoeveelheden van 1000 ml) en koffie (K, standaardhoeveelheden van 250 gr), de door Attwood Panels B.V. verstrekte aankoopgegevens over 1979 uitvoerig geanalyseerd.

Het was Ehrenberg die aandacht vroeg voor de betekenis van het verdelingskarakter (ofwel: het stochastische karakter) van individuele aankoop*gebeurtenissen* en de consequenties daarvan voor het totale aantal aankoopgebeurtenissen in een bepaalde periode t[1]. Hij postuleerde en valideerde een model dat het volgtijdelijke aantal aankoopgebeurtenissen op het individuele niveau beschrijft door middel van een Poisson-verdeling met individuele parameter, λ_i (i.c. het gemiddelde aantal voor huishouding i). Is het nu mogelijk de heterogeniteit over de populatie van gezinnen te beschrijven d.m.v. een gammaverdeling met parameters α en β, dan volgt voor de verdeling van het aantal aankoopgebeurtenissen in een zekere periode (week, maand, e.d.) t een negatief binomiale verdeling met door α en β te beschrijven gemiddelde en variantie[2]. Een andere interessante zienswijze werd ontwikkeld door Morrison. Deze auteur toont aan dat in geval van identiek verdeelde aangekochte *hoeveelheden* per gezin, in periode t, cross-sectioneel onderzoek omvangrijke 'heavy half'-proporties zal opleveren.

Zijn de hoeveelheden bijvoorbeeld identiek normaal verdeeld, met gemiddelde μ en standaardafwijking σ, dan is de proportie $p(hh)$, der totale afzet welke voor rekening komt van de gezinnen die méér dan het gemiddelde μ kopen, gelijk aan $0{,}50 + 0{,}399(\sigma/\mu)$, de zogenaamde 'heavy-half'[3].

Beide bovengenoemde zienswijzen beperken zich tot steeds één specifieke component van het aankoopgedrag. Verhulp introduceerde de gedachte om, in principe, vier stochastische componenten te beschouwen; deze gedachte werd nader uitgewerkt door Van der Zwan en Verhulp[4, 5]. Uitgangspunt is daarbij de volgende definitie van het per gezin i in periode t aan produkt j bestede bedrag:

$$V^j_{t/i} = (X\,Y\,Z\,P)^j_{t/i} \quad , \tag{1}$$

welke blijkens onze ervaringen met frequent gekochte verbruiksgoederen veelal kan worden vereenvoudigd tot:

$$V^j_{t/i} = (Y\,Z\,P)^j_{t/i} \quad . \tag{2}$$

Afhankelijk van de keuze voor (1) of (2) is de volgende toelichting van belang (eenvoudigheidshalve wordt de produktaanduiding j weggelaten):

$V_{t/i}$: de waarde van de door i in t gekochte hoeveelheid; $V_{t/i} \geqslant 0$

$X_{t/i}$: een dummy-variabele welke $= 1$ in geval i in t één of meer keren aankoopt; zo niet, dan geldt $X_{t/i} = 0$;

$Y_{t/i}$: het (voorwaardelijk) aantal aankoopgebeurtenissen door i in t; i.g.v. definitie (1) geldt:
$Y_{t/i} = 1, 2, 3, \ldots$; i.g.v. definitie (2):
$Y_{t/i} = 0, 1, 2, 3, \ldots$;

$Z_{t/i}$: het voorwaardelijk gemiddeld per aankoopgebeurtenis gekochte aantal standaardeenheden door i in t; $Z_{t/i} > 0$.

$P_{t/i}$: de voorwaardelijk gemiddeld per aankoopgebeurtenis betaalde prijs per standaardeenheid door i in t; $P_{t/i} > 0$.

De notatie $V_{t/i}$ etc. duidt op het volgtijdelijke gedrag van gezin i, hetgeen beschreven kan worden met behulp van data uit een gezinspanel. Wij spreken van het *intra*-individuele gedrag, dat als tijdreeks kan worden beschouwd.

Een ander gezichtspunt m.b.t. de diverse gedragscomponenten betreft de verdeling daarvan, in een specifieke periode t, over alle gezinnen. Vergelijk daartoe bijvoorbeeld de eerder genoemde negatief binomiale verdeling der aankoopgebeurtenissen bij Ehrenberg. Wij spreken dan van het zogenaamde *inter*individuele gedrag.

Naar analogie van het 'model' van Ehrenberg hanteren wij vanaf nu een daaraan analoog 'denkschema', gebaseerd op mengvormen ('mixtures'[6]) van verdelingen. Daardoor kan een brug worden geslagen van component verdelingen op het individuele niveau naar relevante aggregaten in periode t. Het geheel van variabelen en verdelingen dat daarbij een rol speelt betitelen wij voortaan als de *aankoopstructuur*.

De uitkomsten m.b.t. het intra-individuele gedrag (verdeling A) worden beschreven in paragraaf 3, die m.b.t. de populatieheterogeniteit (verdelingen B1 en B2) in paragraaf 4 en die t.a.v. het interindividuele gedrag in paragraaf 5.

In paragraaf 6 worden de mogelijkheden besproken om m.b.t. YZ en YZP

Gezinshuishoudingen	Opeenvolgende tijdsperioden 1	2	3	.	.	m	Intra-individuele verdeling (horizontaal)	Paramaters intra-individuele verdeling	
1	X_{11}	X_{12}	X_{13}	.	.	X_{1m}	A_1	P_{11}	P_{12}
2	X_{21}	X_{22}	X_{23}	.	.	X_{2m}	A_2	P_{21}	P_{22}
3	X_{31}	X_{32}	X_{33}	.	.	X_{3m}	A_3	P_{31}	P_{32}
.	.	.	.	.	.	.	.		
.	.	.	.	.	.	.	.		
n	X_{n1}	X_{n2}	X_{n3}	.	.	X_{nm}	A_n	P_{n1}	P_{n2}
Populatieheterogeniteit (verdelingen individuele parameters; verticaal)								B_1	B_2
Interindividuele verdeling (verticaal)	C	C	C	C	C	C			

Figuur 1. Denkschema bij onderzoek

het interindividuele gedrag te simuleren op basis van het intra-individuele gedrag. In paragraaf 7 worden de belangrijkste conclusies samengevat. Eerst wordt nu echter in paragraaf 2 kort ingegaan op de relevantie van dit soort onderzoek.

2. DE AANKOOPSTRUCTUUR: RELEVANTIE VOOR ONDERZOEKOPZET EN BELEID

2.1. *Het marktonderzoek*

Vele vormen en toepassingen van continu en/of ad hoc marktonderzoek leveren in de praktijk een bijdrage aan commercieel-politieke oplossingen. Algemene en/of specifieke beleidsvragen indiceren daarbij voor de ervaren onderzoeker het gebruik van bepaalde onderzoekmodellen, methoden en technieken.

Er bestaan echter ook problemen welke samenhangen met het desbetreffende produktveld en wel in het bijzonder met het aankoopgedrag.

De vraag, welk type steekproef- en registratiemethode ter zake zal moeten worden toegepast, kan in feite niet op bevredigende wijze worden opgelost zonder dat men de beschikking heeft over de nodige kennis omtrent twee hoofddimensies: populatieheterogeniteit en spreiding in de tijd[7].

De eerste dimensie betreft de mate waarin de parameters van het gedrag – bijvoorbeeld: gemiddeld aantal aankoopgebeurtenissen, gemiddeld betaalde prijs, etc. – verschillen tussen de gezinnen.

De tweede dimensie betreft de regelmaat van het individuele aankopen of van de individuele aankoopgebeurtenissen: het maakt een groot verschil of er sprake is van een min of meer vaste regelmaat, van beschrijfbare stochastiek, of van praktisch onbeschrijfbare regelmaat in de handelingen. Is er bijvoorbeeld sprake van (bijna volledige) homogeniteit m.b.t. gedragsparameters en

is er bovendien sprake van een min of meer vaste regelmaat ten aanzien van aankopen of aankoopgebeurtenissen (i.c. alle gezinnen hebben een ongeveer gelijk gemiddelde, èn alle gezinnen kopen ten naaste bij één keer per week, maand, o.d.), dan voldoet een eenmalige enquête bij een beperkte steekproef met een korte registratieperiode.

Vertonen daarentegen de gedragsparameters over de gezinnen aanzienlijke verschillen èn bestaat er aanzienlijke spreiding van de aankopen in de tijd, dan kan alleen panelregistratie (langdurige registratieperiode, grote steekproef), aangevuld met eenmalige registratie op basis van een grote steekproef, inzicht verschaffen ten aanzien van voor het beleid relevante grootheden[8].

Men zou nu kunnen tegenwerpen dat de problematiek bijna niet oplosbaar lijkt, omdat inzicht m.b.t. de actuele niveaus der beide bovengenoemde dimensies pas verkregen kan worden door geruime tijd volgtijdelijk te registreren, dus door gebruik te maken van panelregistratie.

Deze op zich spitsvondige redenering gaat echter voorbij aan het feit dat men in geval van vele praktische toepassingen op grond van a priori informatie (ervaring, reeds geaccumuleerde kennis uit ad hoc onderzoek, vergelijkbare produkten) kan aanvangen met een daarmee samenhangende 'safe' opzet van de registratie.

2.2. *Het (commerciële) beleid*

Beleidsmatig is men traditioneel primair geïnteresseerd in resultaten, zoals het afzetvolume in t, mutaties daarin, de gemiddeld betaalde prijs, etc.[9].

Toch is het onzes inziens noodzakelijk dat ook 'beleidsmakers' in de commerciële sfeer, maar ook in die van de overheid, van consumentenorganisaties en dergelijke, zich rekenschap geven van het feit dat de op geaggregeerd niveau waargenomen realisaties gegenereerd worden binnen aankoopstructuren welke per produkt of tussen produktgroepen (grote) verschillen te zien kunnen geven.

a. *Inherente variabiliteit van kengetallen*

Stochastiek m.b.t. gedragscomponenten op het niveau van de individuele afnemers/gezinshuishoudingen kan in een aantal gevallen aanleiding geven tot variabiliteit in de voor het beleid relevante uitkomsten voor de periode t, $t+1$, $t+2$, etc. Dergelijke gevallen betreffen vooral enigszins extreme waarden voor bepaalde parameters, in samenhang met eindige, niet al te grote populaties.[10]

b. *Marketinginspanningen*

Pogingen van gezamenlijke of individuele aanbieders gericht op blijvende beïnvloeding van het niveau en/of het verdelingskarakter van de afzet, de betaalde prijzen en de omzet, worden gericht op de parameters van het individuele aankoopgedrag. Beschikt men over enige informatie met betrekking tot de effectiviteit in dezen van alternatieve marketingprogramma's, dan kunnen simulatiestudies inzicht verstrekken in de gevolgen voor de interindividuele verdelingen.

c. *Toerekening en signalering*

Mag de aankoopstructuur als (bij benadering) stationair worden beschouwd, hetgeen voor frequent gekochte produkten als door ons bezien op kortere termijn niet onredelijk lijkt, dan mag aan periodefluctuaties van een zekere orde van grootte significante betekenis worden toegekend.

Hiermede bedoelen wij dat dan een relatie mag worden gelegd met bepaalde 'externe' omstandigheden zoals bijvoorbeeld de invloed van promotionele acties, actiereclame, seizoensinvloeden, e.d.

In paragraaf 6 laten wij zien dat het mogelijk is deze '*toerekeningen*' in bepaalde gevallen te maken. Wij doen dat dan binnen een zg. *signaleringssysteem*, waarbij de per periode (14 dagen, 4 weken) gerealiseerde steekproefresultaten voor omzetten en prijzen worden getoetst op het al dan niet significant afwijken van het binnen de geschatte aankoopstructuur te verwachten resultaat. Blijken de verschillen statistisch significant, dan is het verantwoord om de eerder genoemde toerekening aan externe factoren te maken.

Uiteraard dient binnen een dergelijk signaleringssysteem ook aandacht te worden geschonken aan wijzigingen in de spreiding van de componenten.

d. *Prijs-, kwaliteits- en assortimentsbeleid*

In hoofdstuk 14 van zijn boek over 'Het Commercieel Beleid bij Verkoop en Inkoop' ontwikkelde P.J. Verdoorn een interessant denk- en werkmodel t.b.v. de assortimentspolitiek[11].

Op grond van enkele veronderstellingen ten aanzien van cruciale factoren:
- de inkomensverdeling over N gezinnen is bij benadering lognormaal.
- de door gezinnen betaalde prijzen en gekochte hoeveelheden zijn functies van het beschikbare inkomen.

kan worden afgeleid hoeveel gezinnen per prijsinterval verwacht mogen worden, en hoe groot het verwachte afzetvolume aldaar zal bedragen. Empirisch kernpunt daarbij is de cumulatieve verdeling der betaalde prijzen die een logaritmisch normaal karakter zou moeten bezitten.[12] Duidelijke afwijkingen van het logaritmisch normale beeld zouden even zovele indicaties bevatten t.a.v. het commerciële beleid:
- onjuiste tussenruimten tussen opeenvolgende prijzen;
- ontoereikend aanbod bij overigens juist gekozen tussenruimten der prijzen;
- kwaliteitsverschillen zijn niet evenredig met de prijsverschillen.

Naar onze mening verdient het zeker overweging om, in het kader van een systematische analyse van panelgegevens, de prijsverdelingen nauwkeurig te bestuderen[13]. Zowel de mate van aanpassing als het gemiddelde en de standaarddeviatie verschaffen beleidsrelevante informatie.

Later zullen wij zien dat het aanbeveling verdient om naast de prijsverdeling ook die der overige componenten en resultanten (volume, omzet) in de beschouwingen te betrekken.

e. *Segmentatie op grond van aankoopkarakteristieken*

Karakteristieken van het aankoop- en verbruiksgedrag kunnen mogelijkerwijs als classificatie- of segmenteringscriteria worden gebruikt.

Morrison wees op het gebruik van het hierboven in paragraaf 1 beschreven

'heavy-half'-criterium. Het zou de bedoeling kunnen zijn daarmede het segment van de zware verbruikers te identificeren. Wij meldden reeds dat in geval van identiek verdeelde aangekochte hoeveelheden – intra-individueel gedrag – cross-sectioneel onderzoek omvangrijke 'heavy-half'-proporties zal opleveren.

Voor het beleid betekenen dergelijke gevallen echter dat er sprake is van een, qua gemiddeld verbruik, homogene groep.

Het is dan ook noodzakelijk om verdergaand longitudinaal onderzoek op individueel niveau te doen om te kunnen vaststellen of er een beschrijfbare heterogeniteit bestaat, zodanig dat meerdere, qua gemiddelde verschillende, groepen kunnen worden onderscheiden welke op zinvolle wijze aan specifieke marketingprogramma's kunnen worden blootgesteld.

In termen van onze aankoopstructuur betekent het laatste geval dat de verdeling van de gemiddelden over de gezinnen niet gedegenereerd is, maar juist spreiding vertoont.

3. INTRA-INDIVIDUEEL GEDRAG

In de eerste plaats werden de intra-individuele componenten van definitie (2) bezien in hun volgtijdelijke verloop. Per huishouding worden drie frequentieverdelingen vastgelegd:

- de verdeling van $Y_{t/i}$, het aantal aankoopgebeurtenissen;
- de verdeling van $Z_{t/i}$, het gemiddeld per aankoopgebeurtenis gekochte aantal standaardeenheden;
- de verdeling van $P_{t/i}$, de gemiddeld per aankoopgebeurtenis betaalde prijs per standaardeenheid.

Wijnberger rapporteerde reeds eerder over het verdelingskarakter van de aankoopgebeurtenissen op het individuele niveau bij deze drie produkten (fijnwasmiddelen, cola en koffie[14]). Hij toetste de door Ehrenberg gestelde Poisson-hypothese en vond dat, kiezend voor een standaardanalyseperiode van vier weken, deze hypothese (individueel getoetst > 95% betrouwbaarheid) voor de overgrote meerderheid der gezinnen niet verworpen kon worden. Deze uitkomsten worden in onderstaande tabel nogmaals gegeven. Voor de gezinnen waarvoor de Poisson-hypothese wel verworpen diende te worden, constateerde hij een bijzonder regelmatig gedrag m.b.t. de aankoopgebeurtenissen. Hij opperde dan ook de mogelijkheid het gedrag van deze gezinnen te beschrijven met een gedegenereerde verdeling[15].

Gemiddeld gekochte hoeveelheden en prijzen gedragen zich statistisch minder eenduidig, een feit dat samenhangt met voor het aankoopgedrag relevante situationele produkt- en referentiegegevens: men bezoekt altijd dezelfde winkel, men koopt traditioneel één flacon wasmiddel, ouderen vertonen dit beeld in versterkte mate, etc.

De belangrijkste onderscheidende statistische maatstaf blijkt hier de *variantie*. Bij elk produkt vinden wij steeds een (omvangrijke) groep gezinnen waarvoor Z en/of P in het geheel niet, of nauwelijks variëren: wij spreken dan van *invariant* gedrag.

Resteert steeds een groep gezinnen welke wel spreiding ten aanzien van Z en/of P te zien geven. Toetsend op individueel niveau (Kolmogorov-Smirnov toets, 95% betrouwbaarheid) accepteerden wij voor de overgrote meerderheid der gezinnen de nulhypothese van *normale* verdelingen, met gezinsspecifieke gemiddelden en varianties.

Bij de toetsingen wordt steeds uitgegaan van 4-weekse perioden. Alleen voor het aantal aankoopgebeurtenissen is ook een toetsing uitgevoerd voor 1-weekse perioden. Dit teneinde de echte Poisson-kopers te kunnen identificeren.

In de onderstaande tabel vatten wij de onderzoekresultaten m.b.t. de componenten en de resultante van de definitie $V \equiv Y.Z.P.$ samen.

Tabel 1. Typologie van gezinnen op grond van intra-individuele verdelingen (3 produkten)

	Produkt:	FW	C	K
$Y_{t/i}$	– Aantal Poisson-kopers (4-weekse periode)	2.751	3.054	4.101
	– Overigen	0	16	47
$Y_{t/i}$	– Aantal Poisson-kopers (1- èn 4-weekse periode)	2.745	2.812	3.362
	– Overigen	6	258	786
$Z_{t/i}$	– Aantal invariante kopers	1.469	982	938
	– Aantal 'normale' kopers	1.277	2.065	3.085
	– Overigen	5	23	125
$P_{t/i}$	– Aantal invariante kopers	986	915	101
	– Aantal 'normale' kopers	1.758	2.107	4.043
	– Overigen	7	48	4
$V_{t/i}$	– Aantal invariante kopers	953	519	94
	– Aantal 'normale' kopers	1.781	2.543	4.033
	– Overigen	17	8	21

De tabel geeft duidelijk weer dat het intra-individuele element van de aankoopstructuur verschillen laat zien over de componenten. Is het zo dat het aantal aankoopgebeurtenissen, $Y_{t/i}$, als overwegend Poisson verdeeld kan worden aangemerkt, hoeveelheden en prijzen zijn niet overwegend normaal verdeeld. Statistische verwerkingsprogramma's zullen dan ook opties dienen te bevatten voor adequate verwerking van invariante groepen.

Een uit oogpunt van statistische analyse interessante vraag is in hoeverre de componenten op het individuele niveau als onderling onafhankelijk mogen worden beschouwd. Ter beantwoording van deze vraag werden correlatiecoëfficiënten berekend en zodanig verwerkt dat het in tabel 2 volgende globale beeld kan worden gegeven. Berekend werden de correlaties binnen de in totaal 13 4-weekse perioden, waarvan wij hier de (absoluut) hoogste en laagste waarden weergeven. Blijkens de uitkomsten van additionele analyses bleek het midden van de aldus gegeven ranges redelijk overeen te komen met de door ons gezochte gemiddelde correlatie op het individuele niveau.

Tabel 2. Indicatie van de range voor de produkt/moment-correlaties tussen de componenten

Correlatie tussen:	FW	C	K
Y en Z (range)	−0,132 − (−0,03)	0,041 − 0,001	−0,234 − (−0,178)
Z en P (range)	−0,451 − (−0,151)	−0,177 − (−0,088)	−0,211 − (0,097)

Een laatste beschouwing binnen het kader van dit hoofdstuk betreft de huishoudingen wier gedrag niet beschreven kan worden in termen van Poisson- en/of normale verdelingen. Ten aanzien van de 'invariante' versus de 'normale' kopers kunnen geen eenduidige en in het oog lopende verschillen worden opgespoord t.a.v. gemiddelden en spreiding van de componenten en resultanten. Anders ligt het met de echte Poisson-kopers versus de overigen (zie tweede geval voor $Y_{t/i}$ in tabel 1). Over alle drie produkten blijkt er waarlijk sprake van 'contrast'-groepen. De uitkomsten van tabel 3 maken duidelijk dat de niet-Poisson kopers gemiddeld veel meer aankoopgebeurtenissen te zien geven met relatief zeer geringe variantie. Via gekochte hoeveelheden en prijzen, die voor elk produkt specifieke uitkomsten voor de contrastgroepen te zien geven, resulteert een en ander in forse verschillen tussen de gemiddeld bestede bedragen en relatieve standaarddeviaties (VC's).

Deze uitkomsten duiden erop dat er voor cola en koffie belangrijke segmenten van heavy-users bestaan, waarbij de aankoopfrequentie (-gebeurtenissen) als meest bijdragende component kan worden aangemerkt.

Tabel 3. Gemiddelden en spreiding van diverse componenten bij contrastgroepen Poisson en niet-Poisson

	FW				C				K			
	Aankoopgebeurtenissen (4-weekse periode)											
	$\bar{Y}$	SD($\bar{Y}$)	VC	N	$\bar{Y}$	SD($\bar{Y}$)	VC	N	$\bar{Y}$	SD($\bar{Y}$)	VC	N
Poisson-kopers	0,33	0,35	1,06	2.745	0,97	1,11	1,14	2.812	1,59	1,07	0,67	3.362
Overigen	3,31	0,22	0,07	6	3,70	1,34	0,36	258	3,57	0,62	0,17	786
Totaal	0,33	0,37		2.751	1,20	1,36		3.070	1,97	1,26		4.148
	Gekochte hoeveelheid per aankoopgebeurtenis (id.)											
	$\bar{Z}$	SD($\bar{Z}$)	VC	N	$\bar{Z}$	SD($\bar{Z}$)	VC	N	$\bar{Z}$	SD($\bar{Z}$)	VC	N
Poisson-kopers	0,93	0,54	0,58	2.745	2,15	1,58	0,73	2.812	3,14	4,01	1,28	3.362
Overigen	0,94	0,45	0,48	6	3,03	1,77	0,58	258	1,92	0,81	0,42	786
	Betaalde prijs (id.)											
	$\bar{P}$	SD($\bar{P}$)	VC	N	$\bar{P}$	SD($\bar{P}$)	VC	N	$\bar{P}$	SD($\bar{P}$)	VC	N
Poisson-kopers	4,11	1,40	0,34	2.745	1,02	0,32	0,32	2.812	3,40	0,33	0,10	3.362
Overigen	4,61	0,71	0,15	6	0,93	0,19	0,21	258	3,48	0,29	0,08	786
	Besteed bedrag (id.)											
	$\bar{V}$	SD($\bar{V}$)	VC	N	$\bar{V}$	SD($\bar{V}$)	VC	N	$\bar{V}$	SD($\bar{V}$)	VC	N
Poisson-kopers	3,93	2,06	0,52	2.745	3,37	2,94	0,87	2.812	17,29	14,31	0,83	3.362
Overigen	12,88	3,35	0,26	6	10,50	7,29	0,69	258	23,63	10,24	0,43	786

4. POPULATIEHETEROGENITEIT

Het intra-individuele gedrag, zoals beschreven in paragraaf 3, vormt het eerste element van de aankoopstructuur. Het tweede element betreft de (systematiek van) *verschillen* tussen de diverse gezinnen t.a.v. de waarde van de gedragsbepalende parameters: de populatieheterogeniteit.

Uit tabel 1 blijkt dat de componenten worden beheerst door twee parameters, het gemiddelde en de variantie. Wij bespreken nu eerst kort samenvattend de heterogeniteit met betrekking tot de gemiddelden.

a. *Verdelingskarakter van individuele gemiddelden*

Aankoopgebeurtenissen $f(\bar{Y}_i)$

De resultaten in tabel 1 hierboven tonen dat de individuele gemiddelden, $\bar{Y}_i$, in afhankelijkheid van het produkt, in overwegende dan wel zeer grote mate als Poisson-gemiddelden kunnen worden beschouwd.

Tabel 3 toont dat de niet-Poisson kopers relatief zeer geringe spreiding vertonen ten aanzien van hun gemiddelden. Er is dus sprake van primaire heterogeniteit: Poisson-kopers vs overigen.

Secundaire heterogeniteit betreft de verdeling van de gemiddelden binnen deze beide groepen.

Uit de analyses door Wijnberger blijkt dat voor de groep Poisson-kopers ook hier de door Ehrenberg gestelde hypothese, volgens welke de gemiddelden een *gamma*verdeling met parameters α en β bezitten, opgaat[17].

Voor de gevallen waarbij de groep 'overigen' qua omvang van belang is (cola, koffie), blijkt zij bijzonder homogeen te zijn. In tabel 3 hierboven lezen wij af dat het gemiddelde aantal aankoopgebeurtenissen substantieel hoger is dan bij de Poisson-kopers en dat de groep zéér homogeen is, gegeven de variatiecoëfficiënten met waarden van 0,36 resp. 0,17.

In de toekomst zullen wij ons verwerkingsprogramma voorzien van een optie om rekening te houden met een eventueel uitgesproken verdelingskarakter voor de groep 'overigen'. Omdat de spreiding der periodeverdelingen hier gedomineerd wordt door de heterogeniteit bij de Poisson-kopers en het effect van de tussengroepsverschillen, beschouwen wij de groep 'overigen' voorlopig als volkomen homogeen.

Gekochte hoeveelheden $f(\bar{Z}_i)$

Hier zoeken wij naar het verdelingskarakter van de per aankoopgebeurtenis gemiddeld gekochte hoeveelheid, $f(\bar{Z}_i)$ dus.

De verdeling van deze parameter laat zich vrij moeilijk fitten aan een of andere theoretische kansverdeling. De oorzaak ligt vooral in het voorkomen van bepaalde combinaties van discrete verpakkingsgrootten, waardoor geen fraai continuüm voor de hoeveelheidsvariabele ontstaat. Het beste empirische resultaat wordt bij de door ons onderzochte produkten bereikt met behulp van de *lognormale* verdeling.

In onderstaande tabel vindt men het maximale verschil, D(max), dat wordt aangetroffen tussen de werkelijke cumulatieve frequentie en de theoretishe cumulatieve frequentie op grond van momentenschatters.

Tabel 4. Recapitulatie verdelingskarakter $\bar{Z}_i$

Produkt	$\bar{Z}$	Var($\bar{Z}$)	Verd. karakter	D(max)	Parameters logn. verd.
FW	0,9141	0,2173	logn.	0,0711	−0,1917 0,1923
C	2,3696	2,5485	logn.	0,0669	0,7116 0,2657
K	2,4749	5,0052	logn.	0,0679	0,7278 0,2792

Betaalde prijzen $f(\bar{P}_i)$

In tabel 5 recapituleren wij de door middel van lognormale verdelingen beschreven heterogeniteit met betrekking tot de gemiddeld betaalde prijzen. De literatuur m.b.t. prijsacceptatie en minimumprijslijnen verschaft ons aanknooppunten ten aanzien van de relevantie van de lognormale verdeling[18]. Zijn in veel gevallen de per periode betaalde prijzen, bij benadering, lognormaal verdeeld, dan is het logisch dat de parameters die dat gedrag beheersen, i.c. de gemiddelden, lognormaal verdeeld zijn met geringere variantie dan de periodeverdelingen.

Tabel 5. Recapitulatie verdelingskarakter $\bar{P}_i$

Produkt	$\bar{P}$	Var($\bar{P}$)	Verd. karakter	D(max)	Parameters logn. verd.
FW	4,1062	1,0415	logn	0,0449	1,3911 0,0405
C	0,9715	0,0598	logn.	0,0730	−0,0556 0,0506
K	3,4210	0,0932	logn.	0,0395	1,2260 0,0079

De prijsverdelingen tonen een goede aanpassing aan de *lognormale* verdeling, een beeld dat met het hierboven gestelde in overeenstemming is.

Bestede bedragen $f(\bar{V}_i)$

Hoewel het bestede bedrag als resultante van de boven besproken componenten tot stand komt, komt het ons nuttig voor ook de feitelijke heterogeniteit met betrekking tot de gemiddeld per periode bestede bedragen te onderzoeken. Het blijkt dat ook hier de *lognormale* verdeling een goede tot zeer goede fit te zien geeft.

Tabel 6. Recapitulatie verdelingskarakter $\bar{V}_i$.

Produkt	$\bar{V}$	Var($\bar{V}$)	Verd. karakter	D(max)	Parameters logn. verd.
FW	4,2402	4,4699	logn.	0,0239	1,3411 0,2016
C	5,1764	23,2711	logn.	0,0303	1,3498 0,5493
K	18,3979	121,0089	logn.	0,0155	2,7588 0,3106

b. *Verdelingskarakter van individuele spreidingsmaatstaf*

Empirisch bleek de meest bevredigende oplossing te worden gevonden door op het individuele niveau de variatiecoëfficiënten $CV(Z)_i$ en $CV(P)_i$ vast te leggen. Voor de niet invariante gezinnen (tabel 1) bleek heterogeniteit m.b.t. deze variatiecoëfficiënten op bevredigende wijze te kunnen worden beschreven d.m.v. *gamma*verdelingen.

5. INTERINDIVIDUEEL GEDRAG: DE PERIODEVERDELINGEN

In dit hoofdstuk bezien wij de feitelijke gedaanten van de interindividuele periodeverdelingen. Voor het geval van stationair gedrag op het individuele niveau en welbeschreven heterogeniteit met betrekking tot de individuele parameters ligt het verdelingskarakter van de diverse componenten en resultanten voor de perioden 1, ..., t, ..., T vast.

Vanwege de onder 2, 3 en 4 gebleken complexiteit is het niet mogelijk om, zoals bij de partiële analyse door Ehrenberg, op a priori gronden een analytische afleiding te geven voor de mathematische gedaante van de periodeverdelingen.

In de volgende paragraaf zullen wij dan ook simulatie introduceren als hulpmiddel voor het bepalen van de aard van de periodeverdelingen, gegeven de reeds vastgelegde karakteristieken van de aankoopstructuur. Het is wellicht nuttig om hier vast te stellen wat men, onder sterk vereenvoudigde omstandigheden, in het algemeen mag verwachten ten aanzien van het gemiddelde en de variantie der periodeverdeling voor, zeg, component P. Stel dat de door gezin i in een bepaalde periode betaalde prijs normaal verdeeld is:

$$P_i = \bar{P}_i + \varepsilon_i \tag{3}$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{4}$$

Ook de verdeling van $\bar{P}_i$ over alle gezinnen is normaal (populatieheterogeniteit):

$$\bar{P}_i \sim N(\bar{P}, \sigma^2) \tag{5}$$

Deze mengvorm van verdelingen geeft als resultaat voor de verdeling van P over de gezinnen (interindividuele verdeling)[19]:

$$P_i \sim N(\bar{P}, \sigma_\varepsilon^2 + \sigma^2) \quad . \tag{6}$$

Kortom, de interindividuele verdeling heeft hetzelfde verdelingskarakter als de heterogeniteitsverdeling, echter met grotere spreiding ($+\sigma_\varepsilon^2$). Alvorens op de aard van het verdelingskarakter in te gaan, geven wij in tabel 7 hieronder eerst een overzicht van het feitelijke verloop van de gemiddelden van de periodeverdeling voor Y, Z, P en V over de drie produkten. De verschillen tussen de opeenvolgende gemiddelden zijn in het algemeen gering, hetgeen duidt op een situatie van stationariteit. Afwijkingen tussen de gemiddelden dienen o.i. uit de volgende oorzaken verklaard te worden:

i) steekproeffluctuatie; zie bijvoorbeeld de $\bar{P}_t$-reeks voor cola;

ii) steekproeffluctuatie èn één of meer 'tijdelijke verstoringen', welke laatste aan tijdelijke promotionele acties, het seizoen, e.d., dienen te worden toegeschreven; zie bijvoorbeeld de $\bar{P}_t$-reeks voor fijnwas;
iii) steekproeffluctuatie èn (semi-) permanente niveauveranderingen; zie bijvoorbeeld de $\bar{P}_t$-reeks voor koffie.

Deze laatste reeks kan bijvoorbeeld gekarakteriseerd worden als trendmatig dalend tot perioden 233/234, waar een aanloop wordt genomen naar een substantiële prijsstijging in periode 235.

Vanaf periode 235 is de situatie dan stationair.

Verder is het interessant om het *verloop* van de verschillende componenten en hun resultante te vergelijken. In tabel 2 gaven wij reeds indicaties over de produkt/moment-correlaties op het individuele niveau. Uit oogpunt van beleid is het uitermate interessant om na te gaan welke de consequenties zijn voor de samenhang tussen de gemiddelden van de diverse componenten en resultanten in de afzonderlijke perioden.

Fijnwas:

Periode 235 wordt gekenmerkt door ongekende prijsacties, mogelijk als antwoord op de introductie van een nieuw, goedkoop merk. Zowel $\bar{Z}$ als $\bar{Y}$ worden gekenmerkt door forse toenamen. Resultante $\bar{V}$ echter, daalt iets.

De overige mutaties, behalve die m.b.t. $\bar{Y}$ in perioden 233 en 234 waarvoor de zomer waarschijnlijk als verklaring moet worden gezien, kunnen als erratisch worden beschouwd.

Cola:

In de zomer wordt de stationariteit doorbroken door geringere aankoopfrequenties, waardoor ook de omzet een kleine deuk vertoont (let op: het gaat hier om aankopen voor thuisverbruik). In de nazomer volgt herstel naar het normale niveau.

Koffie:

De curve $\bar{P}_t$ is typerend voor een niveauverhoging welke binnen enkele perioden haar beslag krijgt. De prijsmutatie in periode 233 heeft een (speculatieve?) vergroting van $\bar{Z}$ tot gevolg, en zo ook van $\bar{V}$.

Voortgaande prijsstijging in periode 234 leidt dan tot herstel naar de uitgangssituatie voor de volumecomponenten.

N.B.: Let bij de verklaring voor de mutaties in de resultante $\bar{V}_t$ op het feit dat deze gedefinieerd is voor de kopende gezinnen. De mutaties in $\bar{V}_t$ zijn dus niet direct indicatief voor mutaties in de totale omzet, waarbij nl. rekening zou moeten worden gehouden met de proportie kopende gezinnen.

In tabel 7 blijkt bijvoorbeeld dat in periode 233 deze proportie direct reageert op de prijstoename.

Tabel 7. Verloop van de component gemiddelden, besteed bedrag en percentage kopende gezinnen.

Periode	226	227	228	229	230	231	232	233	234	235	236	237	238
Fijnwas													
π_t	27	27	28	28	30	30	30	23	27	32	27	24	20
$\bar{Y}'_t$	1,20	1,23	1,22	1,23	1,24	1,23	1,22	1,17	1,22	1,28	1,24	1,21	1,17
$\bar{Z}_t$	0,89	0,91	0,87	0,91	0,94	0,88	0,89	0,92	0,88	0,99	0,92	0,98	0,90
$\bar{P}_t$	4,28	4,13	4,17	4,20	4,13	4,27	4,31	4,23	4,25	3,50	3,79	4,06	4,18
$\bar{V}_t$	4,17	4,27	4,11	4,26	4,38	4,30	4,32	4,25	4,25	4,10	4,10	4,47	4,16
Cola													
π_t	51	53	55	57	54	54	50	43	53	52	51	50	54
$\bar{Y}'_t$	2,31	2,39	2,37	2,35	2,36	2,36	2,16	2,02	2,28	2,31	2,33	2,32	2,30
$\bar{Z}_t$	2,37	2,30	2,32	2,38	2,37	2,38	2,38	2,37	2,35	2,34	2,34	2,39	2,51
$\bar{P}_t$	0,95	0,95	0,96	0,98	0,98	0,98	1,00	0,98	0,97	0,97	0,97	0,98	0,97
$\bar{V}_t$	5,15	5,24	5,21	5,34	5,32	5,42	4,92	4,53	5,06	5,12	5,21	5,17	5,44
Koffie													
π_t	79	81	80	80	80	81	79	74	72	73	77	80	78
$\bar{Y}'_t$	2,62	2,61	2,60	2,59	2,58	2,57	2,40	2,16	2,35	2,55	2,60	2,62	2,53
$\bar{Z}_t$	2,38	2,42	2,49	2,37	2,42	2,47	2,57	3,10	2,46	2,29	2,34	2,42	2,47
$\bar{P}_t$	3,36	3,25	3,21	3,15	3,11	3,10	3,09	3,27	3,72	3,83	3,82	3,82	3,83
$\bar{V}_t$	18,19	17,80	17,61	16,87	16,72	17,07	16,53	18,69	18,58	19,38	20,23	20,70	21,05

π_t = % kopende gezinnen
$\bar{Y}'_t$ = gemiddeld aantal aankoopgebeurtenissen (kopende gezinnen)
$\bar{Z}_t$ = gemiddelde van het gemiddelde aantal gekochte standaardeenheden per aankoopgebeurtenis
$\bar{P}_t$ = gemiddelde van gemiddeld per aankoopgebeurtenis betaalde prijs per standaardeenheid
$\bar{V}_t$ = gemiddeld besteed bedrag

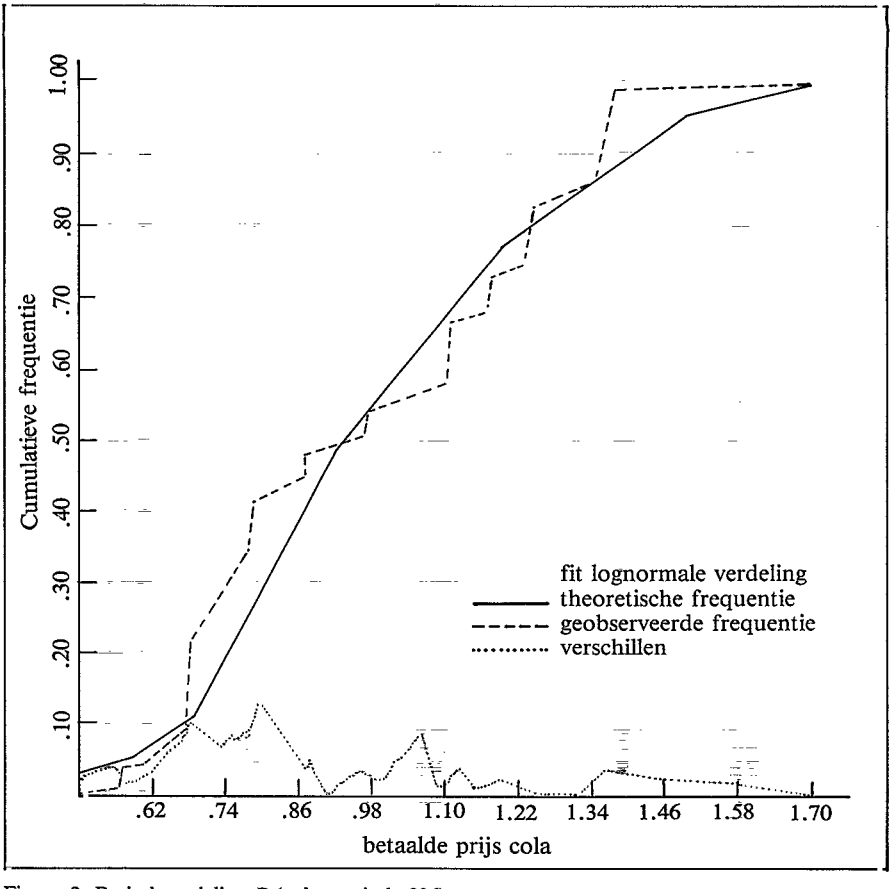
Naast dit inzicht in de *opbouw* van de totale omzet in de loop van de tijd is het van belang het *verdelingskarakter* en de *veranderingen* daarin te bezien.

Gegeven de uitkomsten in paragraaf 4 is het niet verwonderlijk dat de periodeverdelingen voor Z en P als bij benadering *lognormaal* kunnen worden getypeerd.

Evenzomin is het verwonderlijk dat de verdeling voor Y in een periode zeer goed kan worden beschreven door middel van de *negatief binomiale* verdeling (zie paragraaf 1).

Frappant is dat de verdeling voor de bestede bedragen V – ondanks het feit dat met name de verdeling voor de gekochte hoeveelheid Z flinke stappen laat zien – zeer goed kan worden benaderd door een theoretische *lognormale* curve.

Om een indruk te geven van de mate van aanpassing bij de periodeverdelingen geven wij in figuur 2 de waargenomen en de aangepaste lognormale verdeling voor de betaalde prijs van cola (periode 226). De grafiek geeft aan dat er sprake is van een goede fit. De enkele relatief forse afwijkingen worden veroorzaakt door het bestaan van prijsdrempels, zoals bij f 0,69 en f 0,79.



Figuur 2. Periodeverdeling *P* (cola, periode 226).

6. SIMULATIE VAN RESULTATEN, SIGNALERING

Vanwege de beperkingen die aan de lengte van dit artikel gesteld zijn moeten wij voor een uitgebreide beschrijving van dit, met name ook voor de praktijk interessante onderwerp van simulatie en signalering, verwijzen naar het rapport van Verhulp en Wijnberger.[20] Hier zal een zeer beknopte samenvatting worden gegeven.

Indien de parameters en het karakter van de intra- en interindividuele verdelingen zijn vastgelegd, kunnen de periodeverdelingen voor de componenten en resultanten (volume en bestede bedragen) door middel van simulatie worden gegenereerd.

Onder de veronderstelling van stationariteit kunnen deze verdelingen als *referentieverdelingen* worden gekenschetst, analoog aan de negatief binomiale verdeling bij Ehrenberg, die echter analytisch is af te leiden. Een analytische oplossing behoort onzes inziens voor de gehele aankoopstructuur niet tot de mogelijkheden.

De uitkomsten van de simulaties, de referentieverdelingen, vormen een goede weergave van de werkelijkheid, zowel voor de componenten *Y, Z* en *P* als voor de resultante *v*, het volume. Voor de resultante *V* geldt dat er sprake is van een enigszins systematische afwijking.

De meeste afwijkingen echter duiden op wat wij in de meest algemene zin zouden willen benoemen als 'fricties tussen vraag en aanbod'. De referentieverdelingen kunnen onder andere worden gezien als hulpmiddel om dergelijke fricties op te sporen c.q. te signaleren[21].

Daarnaast kan door simulatie inzicht worden verkregen in de gevolgen van zekere systematische veranderingen op het individuele niveau. Bijvoorbeeld kan hier gedacht worden aan de gevolgen van de toenemende vergrijzing van de bevolking in het licht van de nu voor sommige produkten aanwijsbare verschillen tussen het aankoopgedrag van ouderen en jongeren.

7. CONCLUSIES

Het is mogelijk gebleken om aankoopstructuren te kwantificeren. Dat wil zeggen dat door middel van systematische statistische beschrijving de enorme hoeveelheid verschillende individuele handelingen op conciese wijze in kaart is gebracht.

Kennis van de aankoopstructuur voor een produkt maakt het mogelijk om de opzet van ad hoc en continu marktonderzoek te verbeteren. Echter ook ten behoeve van het commerciële beleid is deze kennis van belang.

De mate van aanpassing aan bepaalde statistische verdelingen verschaft in sommige gevallen aanwijzingen voor het assortiments-, kwaliteits- en prijsbeleid. Door middel van simulatie kan worden nagegaan welke de uitwerking zal zijn van bepaalde veranderingen in de bouwstenen van het gedrag van individuele gezinshuishoudingen.

In het desbetreffende onderzoek zijn meer aspecten onderzocht dan er in het kort bestek van dit artikel konden worden vermeld, bijvoorbeeld de samenhang tussen gedragsparameters en bepaalde gezinskarakteristieken.

Zie daartoe ook Verhulp en Wijnberger[10] en een binnenkort te verschijnen volgend rapport, waarin het onderzoek wordt uitgebreid over een groter aantal produkten.

NOTEN

1. Ehrenberg, 1972.
2. Zie bijvoorbeeld de beschrijving van een dergelijke mengvorm van verdelingen: Mood, Graybill en Boes, 1974, p. 123.
3. Morrison, 1968, p. 196.
4. Verhulp, 1977, p. 294.
5. Van der Zwan en Verhulp, 1980, pp. 293-297.
6. Mood, Graybill en Boes, 1974, p. 123.
7. Van der Zwan en Verhulp, 1980, pp. 144-149.
8. Van der Zwan en Verhulp, 1980, tabel 3.21, p. 148.
9. Voor commerciële beleidsmodellen zie bijvoorbeeld Verdoorn, 1964 en Verhulp, 1982.
10. Verhulp en Wijnberger, 1982, werken een aantal casus-posities uit.
11. Zie daartoe enige uitkomsten in hoofdstuk 8.

12. Verdoorn, 1964, pp. 470-473.
13. Van de Woestijne, 1953, hfdst. 6 en 7.
14. Wijnberger, 1980.
15. Wijnberger, 1980, pp. 22-23. Evenzo maakt Bass, 1974, het onderscheid tussen een stochastische en niet-stochastische groep.
16. De χ^2-verdeling en de lognormale verdeling geven minder goede resultaten. Met betrekking tot de toetsing per respondent doet zich ook de vraag voor van de gezamenlijke betrouwbaarheid. Zie hiervoor Wijnberger, 1980, pp. 24-25.
17. Wijnberger, 1980, pp. 35-39.
18. De Koning, 1969, Asselbergs en Van Heteren, 1974, Verhoeff, 1977, Verhulp, 1977 en 1978, Van der Zwan en Verhulp, 1980, pp. 325-333.
19. Raiffa en Schlaifer, 1968, p. 296; Morrison, 1968, p. 197.
20. Zie voor een uitgebreide beschrijving van de simulatieresultaten en de gevolgde werkwijze Verhulp en Wijnberger, 1982, hoofdstuk 7 en appendices 3 en 4.
21. Verhulp en Wijnberger geven als voorbeeld van een dergelijke frictie een gebrek aan aanbod in de lage prijsklasse bij fijnwasmiddelen; Verhulp en Wijnberger, 1982, pp. 39-41.

BIBLIOGRAFIE

Asselbergs, W.J.M. en A.H.J. van Heteren, 1974. *Price Behaviour Models,* opgenomen in: The Challenges facing Marketing Research, How do we meet them? Bundel ESOMAR Congres, Hamburg.

Bass, F.M., februari 1974. *The Theory of Stochastic Preference and Brand Switching,* Journal of Marketing Research, pp. 1-20.

Ehrenberg, A.S.C., 1972. *Repeat Buying, Theory and Applications,* Amsterdam/Londen, Noord-Hollandse Uitgevers Maatschappij.

Koning, C.C.J. de, 1969. *Prijsgedragsmodellen, een Nieuwe Benadering om het Prijsgedrag te bepalen,* in: Marketing Handboek, H.J. Kuhlmeijer e.a., red., Kluwer, Deventer.

Mood, A.M., F.A. Graybill en D.C. Boes, 1974. *Introduction to the Theory of Statistics,* derde druk, McGraw-Hill, Kogakusha, Tokyo.

Morrison, D.G., mei 1968. *A Stochastic Interpretation of the Heavy Half,* Journal of Marketing Research, pp. 194-198.

Raiffa, H. en R. Schlaifer, 1961. *Applied Statistical Decision Theory,* Harvard University Press, Boston.

Verdoorn, P.J., 1964. *Het Commercieel Beleid bij Verkoop en Inkoop,* Stenfert Kroese, Leiden.

Verhoeff, Th.J.M.L., 1977. *De Integraal van de Lognormale Verdeling als Algemene Vorm van de Minimum-Prijslijn,* doctoraalscriptie, EUR, Rotterdam.

Verhulp, J., 1977. *Prijsperceptie: een Empirische Benadering,* Rapport 7702/M, Centrum voor Bedrijfseconomisch Onderzoek, EUR, Rotterdam.

Verhulp, J., 1977. *On Average and Variance of Demand Components in a Purchase Event – Purchase Quantity Model,* in: Modeling for Government and Business, C.A. van Bochove e.a., red., pp. 293-310, Martinus Nijhoff Social Sciences Division, Leiden.

Verhulp, J., 1978. *Prijsacceptatie,* Bedrijfskunde, jaargang 50, 1978/1, pp. 60-64.

Verhulp, J., 1982. *The Commercial Optimum, Theory and Application,* Dissertatie EUR, Rotterdam.

Verhulp J. en J.P. Wijnberger, 1982. *Aankoopstrukturen; verslag van een panelstudie,* Rapport 8203/M, Centrum voor Bedrijfseconomisch Onderzoek, EUR, Rotterdam.

Woestijne, W. van de, 1953. *Een Algemene Vorm van de Vraagfunctie,* Stenfert Kroese, Leiden.

Wijnberger, J.P., 1980. *Statistische analyse van het aankoopgedrag in consumentenpanels; verslag van een onderzoek bij een drietal produkten,* Discussienota 8015/M, Centrum voor Bedrijfseconomisch Onderzoek, EUR, Rotterdam.

Zwan, A. van der, en J. Verhulp, 1980. *Grondslagen en Techniek van de Marktanalyse,* Stenfert Kroese, Leiden.

10. Het consumentengedrag met betrekking tot melk; toepassing van een geïntegreerd consumentengedragsmodel

J.G. TERMORSHUIZEN*

1. INLEIDING

Sinds de jaren zestig zijn enkele pogingen gedaan de processen die te zamen het keuzegedrag van consumenten vormen, weer te geven door middel van zogenaamde geïntegreerde modellen. De meest bekende modellen zijn die van Nicosia (1966), Engel, Kollat en Blackwell (1978) en Howard en Sheth (1969). Deze consumentengedragsmodellen bevatten zowel economische als aan de sociologie en aan de psychologie ontleende variabelen. De modellen, gepresenteerd in de vorm van stroomschema's, inventariseren de variabelen die het keuzegedrag vormen of beïnvloeden en geven aan hoe de auteurs causale relaties veronderstellen aanwezig te zijn.

In dit artikel wordt verslag gedaan van een onderzoek naar het consumentengedrag met betrekking tot melk, dat als doelstelling had meer inzicht te verkrijgen in de gebruiksmogelijkheden van dergelijke consumentengedragsmodellen bij een frequent gekocht consumptiegoed. Deze doelstelling houdt, gezien het produkt waarop ze betrekking heeft, in dat het een onderzoek betreft naar de generieke produktkeuze tussen melk en andere dranken, en dus niet een merkenkeuze. Voorts betreft het een frequent gekocht en geconsumeerd voedingsmiddel. De gevolgen hiervan voor de implementatie van een consumentengedragsmodel worden in dit onderzoek nagegaan.

Het onderzoek heeft betrekking op het produkt consumptiemelk, daar, gezien de afnemende hoofdelijke consumptie ervan en de verschuiving in de vraag van volle naar halfvolle en van gepasteuriseerde naar lang houdbare melk, een groter inzicht in de melkmarkt gewenst was.

Voor een uitgebreide verslaggeving van het gehele onderzoek zij verwezen naar Termorshuizen (1982).

In dit artikel wordt in paragraaf 2 op basis van het model van Engel, Kollat en Blackwell (1978) (verder aangeduid als het EKB-model), een model van het consumentengedrag met betrekking tot melk ontwikkeld. In paragraaf 3 wordt ingegaan op de operationalisering van het model. In paragraaf 4 worden de uitkomsten van enige analyses besproken. Paragraaf 5 bevat de belangrijkste conclusies van het onderzoek.

* Dit artikel is een sterk verkorte weergave van enkele hoofdstukken van het dit jaar verschenen proefschrift van de auteur.

2. MODEL VAN HET CONSUMENTENGEDRAG TEN AANZIEN VAN MELK

In deze paragraaf wordt op basis van het EKB-model een model van het consumentengedrag ten aanzien van melk ontwikkeld. Daaraan voorafgaand wordt het consumentengedrag m.b.t. melk aan een beschouwing onderworpen.

Onze analyse richt zich op de keuze tussen melk en een aantal andere dranken (namelijk: karne- en chocolademelk, volle, magere en drinkyoghurt, vruchtensap, frisdrank, bier, koffie en thee) op het moment van consumptie, dus niet op het aankoopmoment. Hierbij wordt verondersteld dat de dranken in de regel in huis zijn. Melk wordt door de meeste Nederlandse huishoudens jaar in jaar uit min of meer frequent gekocht. Met betrekking tot de keuze van een drank op het ogenblik van consumptie wordt daarom een beperkt probleemoplossend gedrag van toepassing verondersteld.

Het gedrag van consumenten ten aanzien van melk en andere dranken heeft zich in de loop van de tijd gewijzigd: het verbruik van volle melk is afgenomen, dat van koffie, frisdrank en bier is toegenomen; verder zijn nieuwe en vernieuwde dranken, zoals halfvolle melk, drinkyoghurt, karnedrank en vruchtensappen in blokpak geïntroduceerd. Door de reclame via de massamedia en met name bij melk door de veranderde distributiestructuur wordt de consument nadrukkelijk gewezen op de verschillende dranken. De consument is derhalve in zijn koop- en consumptiegedrag niet volledig routinematig. Het beslissingproces wordt van tijd tot tijd opnieuw doorlopen, waarbij men informatie inwint over en zich een beeld vormt van de nieuwe of vernieuwde produkten en ze vergelijkt met bestaande produkten.

Er kan worden aangenomen dat de consument een bepaald beeld heeft van melk. Dit houdt in dat er niet of vrijwel niet actief naar informatie gezocht wordt. Op grond van de in het verleden ontvangen informatie en ervaring vraagt men zich wel geregeld af of men niet eens een andere drank dan de gewoonlijk gekozen drank zal consumeren. De alternatieven worden dus wel telkens, maar niet door iedere consument in even sterke mate, afgewogen.

Centraal in het EKB-model staat een beslissingsproces. In de situatie van uitgebreid probleemoplossend gedrag wordt de consument geacht alle fasen van dit beslissingsproces te doorlopen. De fasen van het model worden genoemd: probleemherkenning, zoeken van alternatieven, evaluatie van alternatieven, keuze. Het resultaat van de keuze is een gevoel van tevredenheid of van ontevredenheid resp. dissonantie. Deze fasen worden nu besproken met het oog op hun relevantie voor een consumentengedragsmodel ten aanzien van melk. Hierbij worden tevens andere dan in het EKB-model opgenomen variabelen aan de orde gesteld, voorzover zij hier relevant lijken.

a. *Probleemherkenning*

Een probleem wordt herkend als de consument een verschil tussen de feitelijke en de ideale situatie voelt dat groot genoeg is om het beslissingsproces in gang te zetten. Aan de ideale situatie liggen bepaalde motieven ten grondslag. Tevens wordt wat als ideaal gepercipieerd wordt, mede bepaald door sociale en culturele factoren.

De aard van de motieven die ten grondslag liggen aan de motivatie van consumenten ten aanzien van melk, wordt in dit onderzoek verondersteld samen te hangen met het moment van de dag of de situatie waarin het probleem herkend wordt. Daarom worden hier vier gebruiksmomenten onderscheiden, nl. ontbijt, tweede broodmaaltijd, loop van de middag en loop van de avond. Het probleem dorst, honger, of anderszins wat de consument voor een drank motiveert, wordt geacht te verschillen met het moment waarop de probleemherkenning zich voordoet. Dit heeft als belangrijk gevolg voor de implementatie dat het model naar een bepaald gebruiksmoment gespecificeerd moet worden.

b. *Zoeken van alternatieven*
Verondersteld wordt dat de consument op de hoogte is van het bestaan en de eigenschappen van de alternatieven. Actief zoeken naar informatie zal zich dan ook niet voordoen, tenzij er sprake is van onlangs geïntroduceerde produktvernieuwingen of produkten.

c. *Evaluatie van alternatieven*
In deze fase worden de relevante alternatieven beoordeeld en vergeleken. De beoordeling geschiedt aan de hand van evaluatiecriteria die in de ogen van het individu van belang zijn. De mate waarin naar de mening van het individu een drank een bepaald evaluatiecriterium bezit, wordt belief genoemd. De oorspronkelijke omschrijving van een belief als de waarschijnlijkheid of onwaarschijnlijkheid dat het object in kwestie gerelateerd is aan een ander object (Fishbein, 1963) is, met het oog op de meting van het model in een landelijke enquête, minder geschikt voor operationalisering.

De beliefs ten aanzien van dranken worden, naar het EKB-model, verondersteld een functie te zijn van normen en waarden, levensstijl en persoonlijkheid. Hetzelfde geldt voor de door de consumenten aan de evaluatiecriteria gehechte subjectieve gewichten. De genoemde verklarende variabelen konden in dit onderzoek niet als zodanig worden gemeten. Zij zijn daarom benaderd door sociale en demografische variabelen. De eerste twee modelvergelijkingen luiden dan ook als volgt:

$$B_{ijk} = f(\text{sociale en demografische variabelen}) \quad (1)$$

$$W_{ikt} = f(\text{sociale en demografische variabelen}) \quad (2)$$

waarin B_{ijk} = belief van individu i m.b.t. evaluatiecriterium k bij drank j.
W_{ikt} = door individu i aan evaluatiecriterium k gehecht gewicht op gebruiksmoment t.

De attitudes worden verondersteld gestructureerd te zijn volgens een expectancy-value model à la Fishbein (1963):

$$AT_{ijt} = \sum_{k=1}^{K} B_{ijk} * W_{ikt} \quad (3)$$

waarin AT_{ijt} = attitude van individu i m.b.t. het drinken van drank j op gebruiksmoment t.
K = het aantal evaluatiecriteria.

d. *Gedragsintentie*

Volgens het EKB-model wordt de gedragsintentie bepaald door de attitude tegenover de alternatieven, sociale invloeden en verwachte omstandigheden. Behalve laatstgenoemde variabele komt dit overeen met Fishbein's gedragsintentiemodel. De sociale invloed in dit model wordt verondersteld aan te geven in hoeverre het individu sociale druk voelt in de richting van een bepaald gedrag en in hoeverre hij aan deze druk toegeeft. Deze subjectieve norm kan op twee manieren opgevat worden, nl. hetzij als de in het algemeen gevoelde druk wat 'men behoort' te doen, hetzij als wat men volgens een of meer bepaalde referentiegroepen behoort te doen (Fishbein, 1967; Ajzen en Fishbein, 1973). Het bij de tweede interpretatie passende model wordt hier toegepast:

$$SN_{ij} = \sum_{m=1}^{M} NB_{imj} * MC_{im} \qquad (4)$$

waarin SN_{ij} = sociale norm van individu i m.b.t. het drinken van drank j.
NB_{imj} = normatieve opvatting van individu i over referentiegroep m m.b.t. het drinken van drank j.
MC_{im} = geneigdheid tot aanpassen van individu i aan referentiegroep m.
M = het aantal referentiegroepen.

De derde variabele met invloed op de gedragsintentie is 'verwachte omstandigheden'. In dit onderzoek is 'de gekozen drank is in huis' de belangrijkste verwachte omstandigheid; hieraan is verder niet expliciet aandacht besteed.

In dit onderzoek wordt de sociale beïnvloeding zoals hiervoor gedefinieerd, beperkt tot referentiegroepen buiten het eigen huishouden. Bij melk zullen ook invloeden binnen het huishouden een niet onbelangrijke rol spelen bij de keuze van een drank door een gezinslid. Als variabele die deze invloed binnen het huishouden weergeeft, is in het onderzoek opgenomen de attitude van een ander gezinslid. Dit is, naast de verbijzondering naar gebruiksmoment, de tweede belangrijke aanpassing van het EKB-model aan het onderhavige beslissingsproces.

Samengevat kan het model van de gedragsintentie ten aanzien van een drank als volgt weergegeven worden:

$$BI_{it} = f(AT_{ijt}, SN_{ij}, AT_{ajt}, AC_t) \qquad (5)$$

waarin BI_{it} = drinkintentie van individu i op gebruiksmoment t.
AT_{ijt}, AT_{ajt} = attitude van individu i resp. a (i en a binnen hetzelfde huishouden) m.b.t. het drinken van alternatief j op gebruiksmoment t.
AC_t = verwachte omstandigheden op gebruiksmoment t.

e. *Keuze*

In het algemeen zal men verwachten dat het alternatief waarvoor men de hoogste gedragsintentie heeft, gekozen zal worden. Er kunnen zich echter omstandigheden voordoen op het moment van de keuze die de beslisser niet verwacht had. In het EKB-model bepaalt de variabele 'onverwachte omstandigheden' samen met de gedragsintentie de uiteindelijke keuze. In dit onderzoek

van het keuzegedrag m.b.t. melk leek het niet in huis hebben van melk de enige relevante en redelijk meetbare onverwachte omstandigheid. De relatie aangaande het gedrag kan als volgt weergegeven worden:

$$C_{it} = f(BI_{it}, UC_t) \tag{6}$$

waarin C_{it} = keuze van individu i op gebruiksmoment t.
BI_{it} = drinkintentie van individu i op gebruiksmoment t.
UC_t = onverwachte omstandigheden op gebruiksmoment t.

f. *Terugkoppeling*
In het EKB-model zijn twee variabelen: dissonantie en tevredenheid opgenomen, die een terugkoppeling naar zoekprocedures resp. het geheugen inhouden. In dit onderzoek is weinig aandacht aan terugkoppelingsvariabelen besteed. Het beslissingsproces komt zo vaak voor dat ervaringen uit vroegere keuzesituaties een stabiel beeld van de onderzochte dranken opgebouwd hebben.

3. OPERATIONALISERING VAN HET MODEL

De in het model opgenomen variabelen zijn gemeten in een mondelinge enquête bij een steekproef van 998 huishoudens. In al deze huishoudens is de huisvrouw geënquêteerd; in 661 huishoudens is tevens aan een a-select gekozen tweede gezinslid een (kortere) vragenlijst voorgelegd, vooral ter meting van AT_{ajt}. Voorzover van toepassing is de vragenlijst gespecificeerd naar een bepaald gebruiksmoment; hierbij zijn onderscheiden ontbijt, tweede broodmaaltijd, loop van de middag en loop van de avond. Iedere respondent is over één gebruiksmoment ondervraagd (split-run). De enquête is uitgevoerd door Interact B.V. te Dongen.

Vergelijking 1
De evaluatiecriteria zijn door perceptie-analyse en open vragen gegenereerd in een vooronderzoek. Als resultaat hiervan zijn zeven attributen naar voren gekomen die mogelijk op één of meer gebruiksmomenten van belang zijn: goedkoop, bijzonder (vs alledaags), verfrissend, gezond, niet slecht voor hart en bloedvaten, lekker, gezellig. Deze attributen komen in grote lijnen overeen met de door Bayton (1963) en Ackerman e.a. (1981) gevonden basisdimensies in de perceptie van melk.

De beliefs t.a.v. de dranken op deze attributen zijn gemeten door middel van vijfpuntsschalen, waarbij de respondenten over iedere van de 84 combinaties van dranken (12) met attributen (7) hun mening moesten geven. De beliefs zijn onafhankelijk van het gebruiksmoment. Het gaat immers om de vraag welke mening men over de dranken zelf, onafhankelijk van een bepaalde situatie, heeft.

Als demografische en sociale variabelen zijn gehanteerd leeftijd, grootte huishouden, sociale klasse, district, opleiding, geslacht, en enkele koopgedragsvariabelen m.b.t. melk zoals distributiepunt, koopfrequentie, voorkeur gepasteuriseerd.

Vergelijkingen 2 en 3
De gebruikssituatie heeft invloed op de aan de attributen gehechte gewichten. Aan de respondenten is daarom gevraagd een gewicht toe te kennen aan de zeven attributen, gegeven het feit dat men op een bepaald gebruiksmoment iets wil drinken. Hierbij is een 70-punts constante-somschaal toegepast. De som van de produkten van gewichten en genormaliseerde beliefs is de attitudescore.

Vergelijking 4
Drie referentiegroepen zijn gespecificeerd: ouders, artsen en kennissen. Voor het meten van de normatieve opvattingen is volstaan met een tweepuntsschaal: eens – oneens. De 'motivation to comply' is via beweringen op een vijfpuntsschaal gemeten. De totale score voor de sociale norm kan hierdoor variëren van nul tot en met vijftien.

Vergelijking 5
De gedragsintentie is gemeten door rangordening van de twaalf dranken. De respondenten moesten zich hierbij voorstellen dat ze 'over een dag of wat' op een bepaald gebruiksmoment iets zouden willen drinken, en vervolgens aangeven welke drank ze in dat geval op de eerste, tweede tot en met de twaalfde plaats zouden kiezen.

Vergelijking 6
Als maatstaf voor het gedrag is de drinkfrequentie van iedere drank op een bepaald gebruiksmoment gehanteerd. De variabele 'onverwachte omstandigheden' is geoperationaliseerd als het aantal maal dat het in een bepaalde periode voorgekomen is dat men melk wilde drinken, maar geen melk kon drinken omdat op dat ogenblik geen melk in huis was.

4. RESULTATEN VAN DE TOEPASSING VAN HET MODEL VOOR DE DRANKENKEUZE

4.1. *Validatie van de evaluatiecriteria*

In deze paragraaf wordt beschreven hoe de respondenten (998 huisvrouwen) melk en andere dranken percipiëren. Deze analyse dient in de eerste plaats ter validatie van de uit het vooronderzoek voortgekomen zeven attributen. Voor deze analyse is de respondenten gevraagd voor telkens twee dranken op een vijfpuntsschaal aan te geven in hoeverre de twee dranken in de ogen van de respondent op elkaar lijken. Analyse van de hieruit resulterende gelijkenismatrix met het MDSCAL-programma (Kruskal en Carmone, 1969) leverde een configuratie van de twaalf dranken in drie dimensies bij 4,5% stress.

De drie dimensies zijn gecorreleerd met de gemiddelden van de scores van de dranken op de zeven attributen, teneinde na te gaan of de perceptuele ruimte in bevredigende mate beschreven kan worden in termen van deze attributen. De hiertoe berekende Spearman-rangcorrelatiecoëfficiënten zijn vermeld in tabel 1.

Tabel 1. Rangcorrelatiecoëfficiënten tussen attributen en perceptuele dimensies

	gezond	alle-daags	ver-fris-send	goed-koop	niet slecht voor 't hart	lek-ker	gezel-lig
dimensie 1	−.26	−.87**	.04	−.53*	−.17	−.52**	−.06
dimensie 2	−.69**	.11	−.15	−.30	−.59*	.11	.83**
dimensie 3	.28	.27	−.69**	.15	−.13	.02	−.31

* $p < .05$; ** $p < .01$

De perceptuele ruimte kan goed worden beschreven door de zeven attributen: ieder attribuut correleert met exact één dimensie significant. Op grond hiervan kunnen de dimensies als volgt beschreven worden. De eerste dimensie maakt onderscheid tussen alledaagse dranken, die tevens niet zo duur en nogal lekker zijn, en minder alledaagse dranken. Koffie en thee zijn in de ogen van de ondervraagde huisvrouwen de meest alledaagse dranken, bier is de minst alledaagse drank. Dimensie twee hangt positief samen met het gezellige karakter van een drank, maar negatief met de beide gezondheidsaspecten. Kennelijk gaan de eigenschappen gezellig en gezond moeilijk samen. Mogelijk hebben bij uitstek gezonde dranken in de ogen van de respondenten een lage 'genotswaarde' en hoeven gezonde dranken ook niet zozeer tevens als gezellig gezien te worden. Deze dimensie lijkt de tegenstelling utiliratistisch (nuttig) tegenover hedonistisch (met hoge genotswaarde) weer te geven. De laatste dimensie verdeelt de dranken in al dan niet verfrissende dranken. Als sterkst verfrissend worden frisdrank, vruchtensap, thee en bier gepercipieerd. De minst verfrissende drank is in de ogen van de respondenten chocolademelk. De positie van volle en halfvolle melk in de perceptuele ruimte is vrij neutraal. Op geen van de dimensies scoort melk extreem. Opvallend is dat er weinig verschil is tussen de perceptie van volle en die van halfvolle melk.

4.2. *Schatting van het model*

Vergelijkingen 1 en 2

Aan de hand van de gewichten die de respondenten aan de zeven attributen hechten is nagegaan of er sprake is van verschillen tussen de gebruiksmomenten (variantie-analyse). De gemiddelde gewichten zijn vermeld in tabel 2.

Tabel 2. Gemiddelden van aan de zeven attributen toegekende gewichten per gebruiksmoment (huisvrouw)

	gezond	bijzonder	verfrissend	goedkoop	niet slecht voor 't hart	lekker	gezellig	totaal
ontbijt	23.4	2.9	6.2	4.4	11.9	13.9	7.3	70
broodmaaltijd	25.7	2.9	7.3	4.0	12.2	11.7	6.2	70
middag	18.1	3.3	9.3	4.1	9.9	13.5	11.7	70
avond	16.9	3.0	8.6	3.2	9.9	13.5	15.0	70
anova: p ≈	.00	.77	.00	.16	.03	.22	.00	
gemiddeld	21.1	3.0	7.8	3.9	11.0	13.2	10.0	70

In vier van de zeven gevallen is sprake van een sterk significant verschil tussen de gemiddelden van de per gebruiksmoment aan de attributen gehechte gewichten. Bij de beide broodmaaltijden let men bij de keuze van een drank meer dan op de overige gebruiksmomenten op gezondheidsaspecten. De gezelligheid en de verfrissende eigenschap van een drank worden juist buiten de broodmaaltijden van groter belang geacht dan bij het brood.

In het algemeen wordt veel gewicht gehecht aan gezondheids- en smaakaspecten (eerstgenoemde met name bij de broodmaaltijden) en, vooral 's middags en 's avonds, aan de gezelligheid. Het lage gewicht van het attribuut goedkoop is mogelijk het gevolg van het feit dat het hier gaat om de keuze van een drank op een gebruiksmoment en niet om een aankoop.

De geconstateerde situationele variabiliteit van de attribuutgewichten werkt door in het gehele beslissingsproces. Immers, de gewichten worden met de beliefs tot attitudescores gecombineerd; de per gebruiksmoment verschillende attitudescores werken op hun beurt door op de gedragsintentie en op het gedrag.

Modelvergelijking 1 is onderzocht met behulp van canonische correlatieanalyse. Hierbij zijn per drank de zeven beliefs de te verklaren variabelen. De in paragraaf 3 genoemde sociale en demografische variabelen zijn de verklarende variabelen. De analyse is voor iedere van de twaalf dranken afzonderlijk uitgevoerd. De uitkomsten worden beknopt besproken.

Het aantal significante ($p < .005$) canonische factoren bedraagt één tot vier. De drie bij volle melk gevonden factoren hebben betrekking op de mate waarin men volle melk goedkoop vindt, op smaakaspecten en op gezondheidsaspecten (can. corr. resp. .35, .30 en .26). Bij halfvolle melk werd slechts één factor gevonden (can. corr. .35) die een op alle punten behalve het gezelligheidsaspect positieve of negatieve houding t.o.v. halfvolle melk uitdrukt. De canonische factoren hangen vooral met opleiding en leeftijd van de respondenten samen. Voor het overige is de samenhang met de verklarende variabelen niet groot.

De tweede vergelijking, die de toegekende gewichten verklaart, is m.b.v. regressie-analyse geanalyseerd voor ieder attribuut en ieder gebruiksmoment afzonderlijk. Van de variantie van de gewichten werd 7 tot 20% verklaard. De belangrijkste verklarende variabelen bleken te zijn opleiding en leeftijd van de respondent en het district.

Vergelijkingen 3 en 4

De attitudescores zijn volgens vergelijking 3 berekend, waarbij de beliefs genormaliseerd zijn. In het algemeen blijkt er sprake te zijn van significante verschillen in de gemiddelde attitudescores m.b.t. de vier gebruiksmomenten. Zowel volle als halfvolle melk scoort tussendoor lager dan bij het brood. Dit verschil wordt veroorzaakt doordat melk hoog scoort op het attribuut gezond, maar uitgesproken laag op het attribuut gezellig, terwijl de gewichten die aan deze twee attributen gehecht worden op tussendoormomenten juist laag resp. hoog zijn.

Halfvolle melk scoort, met uitzondering van 's middags, steeds significant hoger bij de huisvrouwen dan bij de gezinsleden. Volle melk daarentegen wordt door de gezinsleden buiten de broodmaaltijden positiever beoordeeld.

De volgens vergelijking 4 bepaalde sociale norm komt hierna ter sprake.

Vergelijking 5
Deze vergelijking is zowel per respondent als per drank onderzocht. In het eerste geval is voor iedere respondent afzonderlijk onderzocht of er, over de twaalf dranken gezien, sprake is van significante invloed van de verklarende variabelen. In het tweede geval is voor iedere van de twaalf dranken onderzocht of de verklarende variabelen van invloed zijn op de gedragsintentie.

De gedragsintentie is gemeten door rangordening van de twaalf dranken. Zoals ook in ander onderzoek (Ryan en Bonfield, 1975; Bass en Wilkie, 1973) is vergelijking 5 m.b.v. regressie-analyse, een methode die metingen op intervalniveau vereist, geanalyseerd. Opmerkingen omtrent de significantie van de uitkomsten moeten dan ook met enige reserve worden bekeken.

a. *Analyse per respondent*
Vergelijking 5 kan, wegens de verklarende variabele AT_{ajt}, alleen geanalyseerd worden voor de huishoudens waarin naast de huisvrouw een andere persoon ondervraagd is. Voor ieder responderend huishouden waarin dat het geval is, is genoemde vergelijking geschat. In tabel 3 is vermeld voor welk percentage van de respondenten op ieder gebruiksmoment sprake was van significante samenhang tussen de verklarende variabelen en de gedragsintentie.

Tabel 3. Percentages significant ($p<.10$) van nul verschillende regressiecoëfficiënten in het voor iedere respondent in de groep huisvrouwen geschatte model

$$BI_{it} = a + bAT_{ijt} + cSN_{ij} + dAT_{ajt}$$

variabele teken coëfficiënt	AT_{ijt} pos	SN_{ij} pos	AT_{ajt} pos	 neg	n (=100%)
ontbijt	28	28	4	6	127
broodmaaltijd	30	19	2	7	104
middag	31	19	7	6	114
avond	33	11	13	6	113

Bij iets minder dan één op de drie respondenten is er een significante samenhang tussen eigen attitude en intentie. De coëfficiënt van AT_{ajt}, die bij 9 tot 19% van de respondenten significant van nul verschilt, is ongeveer even vaak positief als negatief. Bij de sociale norm is bij 11 tot 28% van de respondenten de coëfficiënt significant positief.

In de loop van de avond heeft voor meer respondenten de attitude van een ander gezinslid een significant positieve invloed op de drinkintentie dan op de overige gebruiksmomenten. De sociale invloed speelt bij het ontbijt bij relatief veel respondenten een rol.

De door de drie opgenomen variabelen verklaarde fractie van de variantie van de gedragsintentie wijkt bij ruim eenderde deel van de geënquêteerde huisvrouwen significant van nul af ($p<.05$), wat overeenkomt met een R^2 van .60 of meer. De mediane waarde van R^2 bedraagt op alle vier gebruiksmomenten omstreeks .50. Deze waarde stemt overeen met in vergelijkbaar onderzoek gevonden waarden.

Tabel 4. Significante coëfficiënten en adjusted R^2 in regressie-analyses van de per drank en per gebruiksmoment geschatte vergelijking 5*

	ONTBIJT				BROODMAALTIJD				MIDDAG				AVOND			
	R^2	AT_{ijt}	SN_{ij}	AT_{ajt}	R^2	AT_{ijt}	SN_{ij}	AT_{ajt}	R^2	AT_{ijt}	SN_{ij}	AT_{ajt}	R^2	AT_{ijt}	SN_{ij}	AT_{ajt}
volle melk	.14	+ +	+ +		.31	+ +			.15	+ +			.18	+ +		
halfvolle melk	.31	+ +			.30	+ +	+ +		.16	+ +			.34	+ +	+ +	+
karnemelk	.04	+ +			.33	+ +			.27	+ +	+ +	+ +	.23	+ +	+ +	
chocolademelk	.05	+ +			.10	+ +			.14	+ +			.10	+ +	+ +	
volle yoghurt	.06	+ +			.23	+ +	+ +		.04	+ +			.06	+ +		
magere yoghurt	.24	+ +	+ +		.13	+ +			.19	+ +	+ +		.11	+ +		− −
drinkyoghurt	.13	+ +		+ +	.16	+ +			.10	+ +			.07	+		+
vruchtensap	.16	+ +			.08	+ +			.08	+ +			.05	+ +		
frisdrank	.06	+	+		.16	+ +			.16	+ +	+ +		.23	+ +	+ +	
koffie	.08	+ +	+ +		.13	+ +		+ +	.11	+ +			.08			+ +
thee	.11	+ +			.23	+ +			.10	+ +			.12	+ +		
bier	.09	+ +			.16	+ +			.04				.27	+ +		− −
n		130				106				115				113		

*+ en −: p<.10; + + en − −: p<.05

b. *Analyse per drank*
De analyse die hiervoor per respondent verricht is, wordt nu per drank uitgevoerd. Het aantal waarnemingen in de regressie-analyse is hier gelijk aan het aantal respondenten in een bepaalde groep. De analyse is voor de twaalf dranken en de vier gebruiksmomenten afzonderlijk uitgevoerd.

In tabel 4 is per gebruiksmoment en per drank aangegeven welke van de in vergelijking 5 opgenomen verklarende variabelen in positieve of negatieve richting significant met de gedragsintentie samenhangen.

De eigen attitude vertoont in 44 van de 48 analyses een sterk significante ($p < .05$) positieve samenhang met de gedragsintentie. De coëfficiënten van de beide andere variabelen zijn in weinig gevallen van betekenis. In het algemeen wordt geen groot deel (nl. 4 tot 34%) van de variantie van de gedragsintentie door de attitudes en de sociale norm verklaard. Bij melk, met name bij halfvolle, ligt dit gunstiger. Met enig voorbehoud kan uit tabel 4 de conclusie worden getrokken dat bij melk, koffie, frisdrank en magere yoghurt sociale invloeden op de meeste gebruiksmomenten van betekenis zijn. De invloed van een ander gezinslid speelt bij de keuze van een drank in de loop van de avond bij meer dranken een positieve of negatieve rol dan op de overige drie gebruiksmomenten.

Vergelijking 6
De omstandigheid dat er geen melk in huis is wanneer men melk wil drinken, bleek zo weinig voor te komen dat daaraan geen aandacht besteed is.

Tussen de gebruiksfrequentie en de gedragsintentie blijkt in de regel een significant positieve relatie te bestaan. De gemiddelde rangcorrelatiecoëfficiënt tussen gebruiksfrequentie en gedragsintentie varieert met de gebruiksmomenten van .51 (tweede broodmaaltijd) tot .57 (loop van de avond).

5. SLOTBESCHOUWING

In dit artikel is op basis van het model van Engel, Blackwell en Kollat (1978) een consumentengedragsmodel ontwikkeld en empirisch getoetst voor een frequent gekocht consumptiegoed. De aard van het produkt, melk en enkele andere alledaagse dranken, en de toepassing van het model op consumptie in plaats van aankoop leidden tot specifieke aanpassingen en uitwerking van het EKB-model. De belangrijkste hiervan zijn:
- het buiten beschouwing laten van inwinning en verwerking van informatie op grond van de aanname dat de consument zich van dergelijke frequent gebruikte produkten een min of meer stabiel beeld gevormd heeft;
- verbijzondering van het model naar gebruiksmoment, op grond van de aanname dat de motieven om te drinken afhankelijk zijn van de situatie;
- verbijzondering van de sociale beïnvloeding naar die van buiten en die binnen het eigen huishouden, op grond van de aanname dat in een consumptiesituatie deze invloed verschillend van karakter kan zijn.

Belangrijke resultaten van de empirische toetsing van het model op basis van een enquête onder 998 huishoudens, zijn de volgende:
- Beliefs vastgesteld door open vragen in een vooronderzoek en door

perceptie-analyse in de definitieve enquête zijn consistent. Er konden zeven beliefs onderscheiden worden, terug te voeren tot drie dimensies. Socio-economische variabelen boden een geringe verklaring van de beliefs.
- Gewichten verschillen duidelijk per gebruiksmoment. Ook hier boden socio-economische variabelen weinig verklaring.
- Attitudes verschillen per gebruiksmoment.
- Sociale invloed binnen het huishouden speelt vooral op tussendoormomenten een rol, invloed van buiten het huishouden vooral bij de keuze van een drank bij het brood.
- De gedragsintentie bleek bij iets minder dan eenderde deel van de respondenten statistisch significant samen te hangen met attitude en sociale invloed.

De analyses leidden tot inzicht in de positie van melk in de ogen van de consument in relatie tot de positie van andere dranken. De positie van melk is vooral sterk bij de tweede broodmaaltijd. Op tussendoormomenten is de positie van melk echter zodanig dat melk niet snel een op dergelijke momenten veel gekozen drank zal zijn. Uit dit onderzoek is gebleken dat het EKB-model ruimere toepassingsmogelijkheden heeft dan alleen voor merkenkeuzeproblemen bij een bepaald produkt. Met de nodige aanpassingen aan specifieke probleemstellingen is het model toepasbaar bij het onderzoek naar beslissingsprocessen van consumenten op vele terreinen.

Naast de compositionele benadering is op de gedragsintentie een andere benadering toegepast, nl. volgens het decompositionele vectormodel. Voorts zijn buiten het in dit artikel aan de orde gestelde gedeelte van het onderzoek andere facetten van het consumentengedrag met betrekking tot melk onderzocht, zoals de ontwikkeling in de loop van de tijd van de voorkeuren van de consumenten en smaakvoorkeuren. Voor de uitkomsten hiervan wordt verwezen naar het proefschrift van de auteur van dit artikel.

REFERENTIES

Ackerman, L., K. Callier en G. van Heghe, 1981. *Konsumentenonderzoek op de Markt van Konsumptiemelk,* LEI-schriften nr. 220/RR-179, Landbouw-Economisch Instituut, Brussel.

Ajzen, I. en M. Fishbein, 1973. *Attitudinal and normative variables as predictors of specific behaviors,* Journal of Personality and Social Psychology, vol. 27, p. 41–57.

Bass, F.M. en W.L. Wilkie, 1973. *A comparative analysis of attitudinal predictions of brand preference,* Journal of Marketing Research, vol. 10, p. 262–269.

Bayton, J.A., 1963. *Contributions of psychology to the microeconomic analysis of consumer demand for food,* American Journal of Agricultural Economics, vol. 45, p. 1430–1437.

Engel, J.F., R.D. Blackwell en D.T. Kollat, 1978. *Consumer Behavior,* The Dryden Press, Hinsdale.

Fishbein, M., 1963. *An investigation of the relationships between beliefs about an object and the attitude toward that object,* Human Relations, vol. 16, p. 233–239.

Fishbein, M., 1967. *Readings in Attitude Theory and Measurement,* Wiley, New York.

Howard, J.A. en J.N. Sheth, 1969. *The Theory of Buyer Behavior,* Wiley, New York.

Kruskal, J.B. en F.J. Carmone, 1969. *How to use MDScal and other useful information,* Bell Laboratories, Murray Hill.

Nicosia, F.M., 1966. *Consumer Decision Processes,* Prentice Hall, Englewood Cliffs.

Ryan, M.J. en E.H. Bonfield, 1975. *The Fishbein Extended Model and consumer behavior,* Journal of Consumer Research, vol. 2, p. 118–136.

Termorshuizen, J.G., 1982. *Het Consumentengedrag met betrekking tot Melk,* proefschrift Wageningen.

11. Waargenomen risico

PH. P. TODD

SAMENVATTING

'Waargenomen risico' (Eng.: Perceived risk) kan beschreven worden als de hoeveelheid conflict of spanning die een produkt kan oproepen in de koopsituatie.

Naar het begrip waargenomen risico is met name rond het begin van de zeventiger jaren veel onderzoek verricht, naar aanleiding van de introductie van dit begrip door een team van sociaal-psychologen aan de Harvard universiteit.

Waargenomen risico zou in een aantal nader te omschrijven situaties een belangrijke, remmende invloed hebben op het koopgedrag, zodat grote behoefte ontstond naar nader onderzoek (niet in de laatste plaats vanuit de producenten van consumentengoederen).

In dit literatuuroverzicht zullen wij de belangrijkste elementen de revue laten passeren, die betrekking hebben op waargenomen risico, zoals definities, meetmethoden en manieren om risico te verkleinen.

1. HET BEGRIP 'WAARGENOMEN RISICO' NADER UITGEWERKT

De statistische waarschijnlijkheid dat iemand door een auto zal worden aangereden als hij de straat oversteekt, kan berekend worden met behulp van statistische kansberekeningen. Dit noemen we het 'kansaspect' van risico, ofwel de 'werkelijke' kans. Vergelijkbare voorbeelden zijn de statistische kansen om de voetbaltoto te winnen of om een vermogen te verliezen in het Casino.

Dit kansaspect van risico wordt meestal door statistici bestudeerd: het heeft het voordeel dat de uitkomsten duidelijk en ondubbelzinnig zijn.

In de sociale wetenschappen echter wordt verondersteld dat zowel de situatie als de persoon invloed uitoefenen op het gedrag. Dientengevolge zal men bij bestudering van het consumentengedrag niet zozeer geïnteresseerd zijn in de werkelijke kans op risico, als wel in de *perceptie van* en de *reacties op* risico's. En dat is, omdat de consument alleen *kan* reageren op zijn/haar subjectief waargenomen en geïnterpreteerde risico-taxatie van een bepaalde (koop)situatie. Alleen de perceptie van dat risico kan het koopgedrag of de koopintentie beïnvloeden. Uit de sociaal-psychologische literatuur is bekend dat er in een bepaalde situatie sprake is van een grote *over-*, respectievelijk *onderschatting* van het 'werkelijke', statistische risico dat men loopt (zie o.m. Stallen, 1979 en Vlek & Stallen, 1980).

Een bekend voorbeeld is het feit dat veel mensen vliegen riskanter vinden dan autorijden, terwijl dit laatste statistisch gezien vele malen riskanter is.

In dit overzicht zullen we ons verder uitsluitend bezighouden met *waargenomen* risico, dus het risico zoals dat wordt waargenomen en onder woorden gebracht door consumenten.*)

* In dit verband gaan wij derhalve voorbij aan de (ruime) hoeveelheid sociaal-psychologische literatuur omtrent risico*name,* zoals die door Kogan en Wallach (1964, 1967) e.v.a. is bestudeerd.

Bauer (1960) was de eerste die het consumentengedrag beschouwde als een vorm van risiconame. Hij stelde dat:

'Consumentengedrag bepaalde risico's impliceert, in die zin dat elke actie van de consument bepaalde gevolgen met zich mee zal brengen die niet (alle) te voorzien zijn met 100% zekerheid'.

En dat:

'Consumenten karakteristieke besluitvormingsstrategieën ontwikkelen, alsook manieren om risico's te verkleinen. Dit stelt hen in staat om met redelijk vertrouwen en op een gemakkelijke manier te handelen in situaties waarbij ze onvoldoende informatie hebben en waarbij de gevolgen van hun handelen redelijkerwijze niet zijn na te gaan'. (Op. cit. p. 24, 25.)

De in dit citaat genoemde besluitvormingsstrategieën duiden erop, dat Bauer in feite uitgaat van de zg. *cognitief-rationele* benadering van het consumentengedrag: Daarbij is het centrale probleem voor de consument de *keuze* tussen aankoopalternatieven (zie Hansen, 1972).

Bauer stelt dus, dat elk soort consumentengedrag een zekere mate van waargenomen risico met zich meebrengt dat niet altijd voorspelbaar is (*onzekerheid*) en dat wel eens onaangename consequenties zou kunnen hebben (*gevolgen*).

Tijdens de afgelopen twintig jaar is er veel onderzoek verricht en gepubliceerd met als onderwerp het begrip waargenomen risico. Stem *et al.* (1977) geven een aardige indeling in soorten van onderzoek die zijn verricht; in figuur 1 hebben wij een bewerkt en aangepast raamwerk weergegeven, waarin alle onderzoeksterreinen kunnen worden geplaatst.

We zien dat waargenomen risico vanuit drie invalshoeken benaderd kan worden: vanuit de socio-economische en vanuit de situatieve invalshoek en vanuit de invalshoek van de persoonlijkheid. Het begrip waargenomen risico zelf kan omschreven worden in termen van *niveaus* (macro-, meso-, micro-), in termen van *componenten* en in termen van *soorten*.

Een apart aandachtsgebied omvat de verschillende *meetmethoden* van waargenomen risico, omdat er rond dit punt nog veel controversen bestaan.

Het model mondt uit in het resulterende gedrag na de risicoperceptie en -verwerking, t.w. de *risicobehandeling,* onder te verdelen in risico*reducerend* en risico*vergrotend* gedrag.

Waargenomen risico leidt niet automatisch tot risicobehandeling: het kan ook in latente vorm aanwezig zijn.

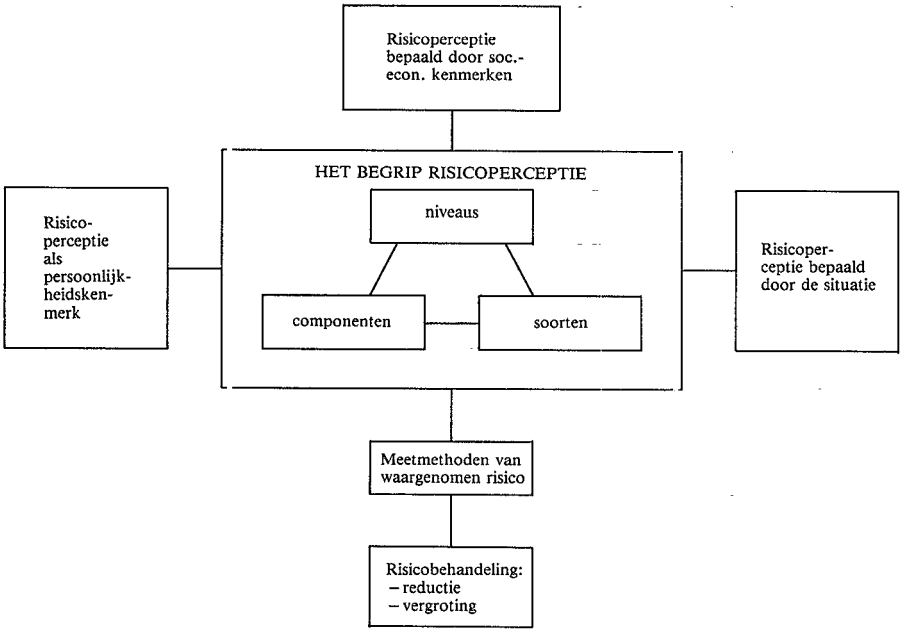
Er zijn een aantal voorwaarden (waarover later meer), waaronder waargenomen risico resulteert in risicobehandeling.

Anderzijds zien we dat men de aanwezigheid van waargenomen risico veronderstelt wanneer er risicobehandeling optreedt.

Na een bespreking van het begrip waargenomen risico zal een overzicht worden gegeven van de niveaus, de componenten en de soorten risico, zoals deze in de literatuur voorkomen.

Een goede definiëring van waargenomen risico is van belang, omdat daar in feite de operationalisatie (lees: meetmethode) van afgeleid wordt.

Ondanks de grote hoeveelheid onderzoek die verricht is op dit terrein, bestaat er tot op de huidige dag nog geen overeenstemming omtrent een duidelijke



Figuur 1. Een raamwerk voor de analyse van risicoperceptie.

conceptuele *definitie* van het begrip waargenomen risico. Onzes inziens is dat de belangrijkste reden dat momenteel dit onderwerp aan belangstelling heeft ingeboet. Hieronder volgt een kleine bloemlezing van bestaande definities.

Taylor (1974) beschouwt waargenomen risico als een equivalent van onzekerheid, terwijl Peter en Ryan (1976) spreken over 'de verwachte verliezen, verbonden aan de aankoop van een produkt'. Beide definities zijn o.i. te beperkt.

Spence *et al.* (1970) definiëren waargenomen risico in operationele termen als 'de hoeveelheid risico die een respondent zegt waar te nemen bij de specifieke aankoopsituatie van een produkt'. De hoeveelheid risico werd dus 'direct' gemeten door middel van de volgende vraag: 'Hoeveel risico ervaart u bij de aankoop van...' Het is duidelijk dat op deze wijze de interpretatie van het begrip risico bij de respondent wordt gelegd...

Bettman (1973, 1975) geeft geen algemene definitie, doch suggereert dat er twee componenten te onderscheiden zijn, t.w. 'inherent' en 'gehanteerd' risico. Waargenomen risico is aldus een combinatie van 'de ingebakken hoeveelheid conflict die een bepaalde produktklasse teweeg kan brengen' en 'de hoeveelheid conflict die (nog) resteert nadat informatievergaring en risicoreductie hebben plaatsgevonden'. (Op. cit. 1973, p. 184.)

Stallen (1979) definieert risico als de 'waargenomen mogelijke schade en/of verlies van een keuze-alternatief'.

In het woord 'mogelijke' komt de mate van onzekerheid tot uitdrukking, terwijl in het aspect 'schade/verlies' de subjectieve waardering van ongewenste gevolgen van een keuze-alternatief is begrepen.

Volgens Cox (1967a) zal het vaak noodzakelijk zijn de aanwezigheid van

waargenomen risico *af te leiden* uit een aantal signalen, omdat '...consumenten wellicht niet in staat zijn of niet willen toegeven dat een situatie waarmee ze geconfronteerd worden, riskant is... (en daarom) ...gaan we ervan uit dat er sprake is van risicoperceptie wanneer men overgaat tot risicobehandeling, zoals bv. risicoreductie'. (Op. cit. p. 36, 37.)

De aanwezigheid van waargenomen risico wordt aldus afgeleid uit de gevolgen ervan, zoals risicobehandeling etc.

We komen hier op het terrein van de meetmethoden van risicoperceptie (zie verder paragraaf 4).

Naar onze mening komen de definities van Bettman en van Stallen het dichtst bij een *theoretische definitie,* die wij als volgt willen omschrijven:

Waargenomen risico is het algemene gevoel van onbehagen dat ontstaat bij het waarnemen van mogelijke schade en/of verlies van een keuze -alternatief. De schade en het verlies kunnen van materiële of van immateriële aard zijn.

Wat wij onder het 'algemene gevoel van onbehagen' verstaan, wordt door Howard en Sheth (1969) de zg. stimulus ambiguïteit genoemd in hun bekende model van het consumentengedrag.

Meestal leidt dit onbehagen tot de drang ('arousal') om informatie te gaan verzamelen.

Onder 'keuze-alternatief' kan zowel een produktklasse (bv. wasmachines) als een specifiek merk/model worden begrepen, afhankelijk van de situatie.

2. COMPONENTEN, NIVEAUS EN SOORTEN VAN RISICOPERCEPTIE

Omdat de meeste onderzoekers op het gebied van risicoperceptie zich op steeds wisselende niveaus, componenten en soorten van risico's hebben gericht, willen we hierna op deze begrippen wat nader ingaan.

Daarbij gaan we uit van de aankoopsituatie waarin een consument kan verkeren: welke componenten, niveaus en soorten van risico neemt hij/zij waar bij de aankoop van produkten of diensten?

2.1. *Componenten*

De meeste onderzoekers gaan uit van twee componenten van risicoperceptie, nl. 'onzekerheid' en 'gevolgen'.

Cox (1967a) stelt dat de hoeveelheid waargenomen risico een functie is van: a) 'datgene wat verloren zou gaan of wat *op het spel staat,* wanneer de uitkomsten van de handeling ongunstig zouden zijn, en b) het *subjectieve gevoel van zekerheid* dat de gevolgen ongunstig zullen zijn.' (Op. cit. p. 37.)

Datgene wat op het spel staat is een functie van het belang van de (koop)-doeleinden, de zwaarte van de 'straf' die staat op het niet bereiken van die doelen, en, niet in het minst, de hoeveelheid middelen (lees: geld) die aangewend zijn om de doeleinden te bereiken.

Het zal u duidelijk zijn dat de hoeveelheid risico alleen verminderd kan worden wanneer een van de twee of beide risicocomponenten worden gereduceerd, bv. door middel van informatievergaring. Wij komen hierop terug in een latere paragraaf.

Deze twee-componententheorie van risicoperceptie is met name verspreid

door de eerder vermelde Harvard psychologen (Cox (1967), Cox en Rich (1964), Cunningham (1967a)). Anderen gaan uit van alleen onzekerheid (Arndt (1967a)), of van onzekerheid, bestaande uit: 1. onzekerheid omtrent de resultaten en 2. onzekerheid omtrent de gevolgen (Taylor (1974)).

Gezien het voorgaande, en in het licht van onze definitie, lijken Arndt en Taylor een te beperkte benadering te hanteren.

Een geheel andere benadering is gekozen door Bettman (1973).

Hij maakt onderscheid tussen twee componenten, nl.
a) inherent risico en b) gehanteerd risico.

Hij definieert *inherent risico* als het 'latente' risico dat een produktklasse in zich heeft voor een consument. Het is de 'ingebakken' hoeveelheid conflict die een bepaalde produktklasse teweeg kan brengen. Dus een soort intuïtief, emotioneel gevoel ten opzichte van de produktklasse.

We zien dat in feite alle studies van de Harvard groep (zie Cox (1967)) zich beperken tot dit inherente risico.

Gehanteerd risico is de hoeveelheid conflict die een aankoop veroorzaakt wanneer de consument een speciaal merk of type kiest in een bepaalde koopsituatie.

Gehanteerd risico vertegenwoordigt het eindresultaat van informatievergaring en risicoreductieprocessen die inwerken op het inherente risico.

Terwijl er bij inherent risico sprake is van affectieve elementen (intuïtie, emotie), zien we dat bij het gehanteerde risico meer de cognitief-rationele elementen optreden.

Bettman veronderstelt de volgende relatie tussen beide risicocomponenten: Wanneer de consument in 't geheel geen informatie bezit omtrent de produktklasse, dan hebben beide risicocomponenten dezelfde waarde (laag of hoog). Wanneer evenwel de hoeveelheid verworven (nuttige) informatie groter wordt, zal het gehanteerde risico kleiner worden.

In theorie is dit onderscheid in beide risicocomponenten plausibel; in de praktijk hebben wij nogal wat bezwaren tegen de operationalisaties van Bettman (zie paragraaf 3), waardoor deze componenten minder bruikbaar zullen zijn in de praktijk van het marktonderzoek.

2.2. *Niveaus*

We kunnen volgens Cox (1967b) drie risiconiveaus onderscheiden, te weten een *specifiek,* een *tussenliggend* en een *algemeen* risiconiveau. Het specifieke niveau heeft betrekking op bv. de aankoop van een bepaald merk van de produktklasse auto's, het tussenliggende niveau op de produktklasse zèlf (auto's kopen), en het algemene niveau op een gehele groep van situaties, bv. gebruikspatronen of aankoopgewoontes.

Andere voorbeelden van produktvelden en respectieve risiconiveaus (van algemeen naar specifiek) zijn: Voedingsmiddelen - diepvrieswaar - Iglo spinazie. Financiële zaken - leengedrag - Amro Continu Crediet.

De meeste onderzoeken van risicoperceptie hebben betrekking op het specifieke en het tussenliggende niveau. Het blijkt nl. dat op deze niveaus de verbanden met het consumentengedrag sterker zijn dan op het algemene niveau (zie Hisrich et al. 1972).

2.3. *Soorten*

De hierna te bespreken soorten van risicoperceptie verwijzen naar de verschillende risico-*inhouden*. Jacoby en Kaplan (1972) vonden zes risicosoorten bij een door hen gehouden literatuuronderzoek, nl.:

1. Financieel risico:
Wanneer een ons onbekend produkt kapot gaat of dienst weigert, is ons verlies het geld dat het kost om het te laten repareren of het te vervangen.

2. Prestatierisico:
De waargenomen waarschijnlijkheid dat er iets mis zal zijn met een onbekend merk of dat dit merk niet goed zal functioneren (bv. nieuwe diaprojector weigert).

3. Fysiek risico:
Sommige produkten zijn gevaarlijk voor de gezondheid of de veiligheid, wanneer ze ondeugdelijk zijn (denk aan slechte isolatie van een electrisch broodrooster).

4. Psychologisch risico:
Het gevaar dat een onbekend merk niet past bij de manier waarop we over onszelf denken (= ons zelfconcept). Een voorbeeld kan zijn een door ons gekocht (nieuw) merk sigaretten: de smaak is goed, doch het advertentiemodel op de reclameposters staat ons niet aan.

5. Sociaal risico:
Het gevaar dat we onszelf belachelijk zullen maken tegenover bekenden, wanneer het produkt een sof blijkt te zijn (bv. serveren van taaie biefstuk door gastvrouw aan nieuwe chef van echtgenoot).

6. Risico van tijdverlies:
Wanneer produkten weigeren, dan kost ons dat meestal tijd, moeite en ongemak om de zaak recht te zetten (bv. 'maandagmorgen'-auto met steeds nieuwe kwaaltjes).

Sommige auteurs maken geen onderscheid tussen (4) en (5), en spreken van een *psycho-sociaal* risico (Cox, 1967) of van een risico van 'Ego verlies' (Roselius, 1971).

Resumerend zien we dat in bovenstaande zes risicotypen in feite drie basissoorten voorkomen, t.w.:
- Risico's met betrekking tot geld/tijd/moeite (1 en 6)
- Risico's met betrekking tot het (gekochte) object (2 en 3)
- Risico's met betrekking tot het (kopende) subject (4 en 5).

Het hangt van de produktklasse af die we bestuderen, welke soort risico dominant zal zijn over de andere.

In de eerder genoemde studie van Jacoby en Kaplan werden zowel alle risicosoorten (m.u.v. risico van tijdverlies) alsook een algemene risicomaat ge-

hanteerd bij twaalf verschillende consumentenprodukten. Het bleek dat bij acht van de twaalf produkten het prestatierisico het sterkst correleerde met de algemene risicomaat. Daarna volgden (in aflopende rangorde): financieel, sociaal, psychologisch en fysiek risico.
Als algemene risicomaat hanteerden de auteurs de antwoorden op de volgende vraag: 'On the whole, about how risky would you say it was to buy an unfamiliar brand of ... (name product)?'

Dit is dus een directe vraagstelling naar de risicobeleving van de respondent, waarbij de interpretatie van het woord 'risky' geheel wordt overgelaten aan de ondervraagden.

In het algemeen, en zeker bij meting van waargenomen risico's, denken wij dat meer verfijnde meetmethoden de voorkeur verdienen (zie verder paragraaf 3).

Alle risicosoorten die we hierboven lieten zien kunnen ingebed worden in de beide eerder genoemde risicocomponenten.

Peter en Tarpey (1975) hebben zowel soorten als componenten in het volgende heldere 'expectancy-value' model samengevat:

$$OPR_j = f \sum_{i=1}^{n} (PL_{ij} \cdot IL_{ij})$$

waarbij
OPR_j = totale hoeveelheid waargenomen risico voor merk j
PL_{ij} = kans op verlies i door aankoop merk j
IL_{ij} = belangrijkheid van verlies i bij aankoop van merk j
n = risicosoorten.

Op een inzichtelijke wijze laat het model zien hoe alle risicovariabelen met elkaar samenhangen, en resulteren in het totale, waargenomen risico.

Uit de specificatie van dit model valt op te maken dat Peter en Tarpey het hebben over het *specifieke* risiconiveau, zoals behandeld in paragraaf 2.2. Echter, wanneer we 'merk j' vervangen door 'produktklasse j' of door 'gebruiksgewoonte j', dan omvat het model ook het tussenliggende en algemene risiconiveau.

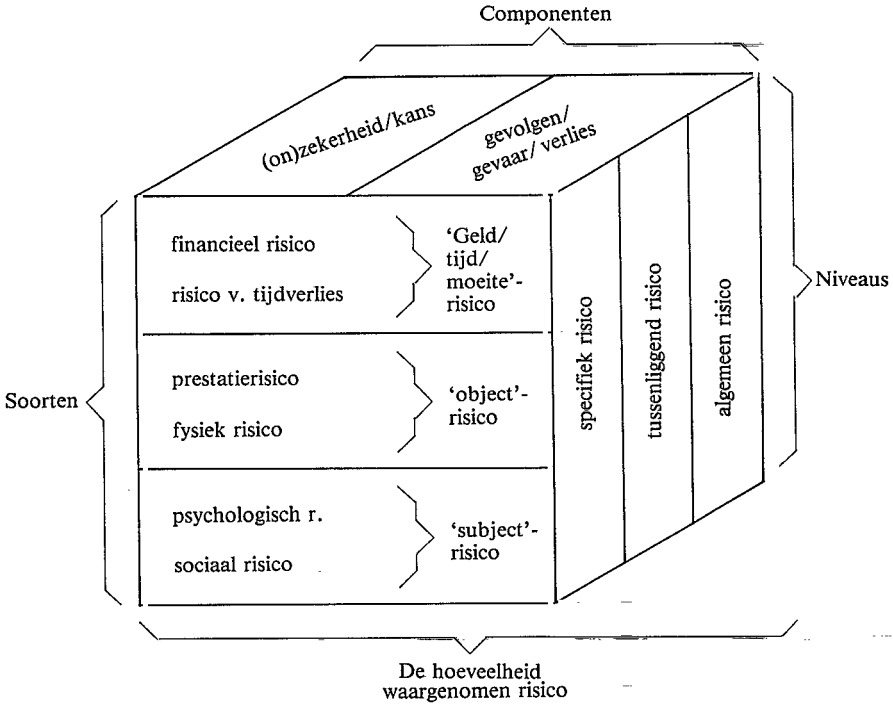
Na bespreking van de belangrijkste risico-elementen, kunnen we een en ander ook ruimtelijk weergeven in een driedimensionaal model, waardoor de onderlinge samenhang nog eens wordt benadrukt (zie figuur 2).

Alle elementen te zamen dragen bij tot de hoeveelheid waargenomen risico die men op een bepaald moment zal ervaren.

In de volgende paragraaf gaan wij dieper in op de meetproblemen die aan dit concept verbonden zijn.

3. HET METEN VAN WAARGENOMEN RISICO

Zoals wij reeds eerder hebben geconstateerd, is het begrip 'waargenomen risico' nooit in theoretische zin gedefinieerd in de onderzoekliteratuur. Het gevolg daarvan is dat de nadruk kwam te liggen op de operationele definities, met andere woorden, hoe een en ander gemeten zou moeten worden.



Figuur 2. Een driedimensionale weergave van componenten, niveaus en soorten van risicoperceptie, resulterend in de hoeveelheid waargenomen risico.

Vlek en Stallen (op. cit.) onderscheiden drie manieren om risicoperceptie te meten:
1. *fysiologische* metingen, zoals hartslag, huidweerstand, en spierspanning;
2. *gedrags*metingen/-observaties, zoals winkelgedrag, etc.;
3. *cognitieve* meetmethoden, zoals verbale omschrijvingen, opinies, uitgesproken voorkeuren, etc.

In de door ons onderzochte literatuur wordt verreweg het meest gebruik gemaakt van de derde meetmethode; het grote probleem daarbij is dat de mate van waargenomen risico sterk wordt bepaald door de *wijze* van vraagstelling, of de *vaardigheid* van de respondent om zijn gevoelens te verwoorden.

Een voor de hand liggende oplossing van dit probleem zou zijn om risicoperceptie op meer dan één van de drie aangegeven manieren te meten, hetgeen echter zelden of nooit wordt gedaan.

3.1. *Vaak gehanteerde meetmethoden*

De meest gebruikte methode om waargenomen risico te meten is de operationalisatie van Cunningham (1967a), die bestaat uit twee componenten, nl. zekerheid en gevolgen.

Om 'zekerheid' te meten hanteerde hij de volgende vraagstelling:*

'Would you say that you are: very certain, usually certain, sometimes certain, or almost never certain that a brand of... (product)... you haven't tried will work as well as your present brand?'

'Gevolgen' mat hij als volgt:

'Wie all know that not all products work as well as others. Compared with other products, would you say that there is: a great deal of danger, some danger, not much danger, or no danger in trying a brand of... (product)... you never used before?' (Op. cit. p. 84.)

De antwoorden op deze vragen werden vervolgens vermenigvuldigd, waarna verschillende risicogradaties ontstonden (zie tabel 1).

Tabel 1. Constructie van de risicoperceptieschaal van Cunningham

	Matrix van waargenomen risico		
	Gevolgenschaal ('danger'):		
	A great deal of danger	Some/not very much danger	No danger
gewicht →	(1)	(2,5)	(4)
Zekerheidschaal: ↓			
Very certain	(4) 4*)	10	16
Usually/ sometimes c.	(2,5) 2,5	6,25	10
Almost never certain	(1) 1	2,5	4

	Risicoperceptieschaal afgeleid uit matrix		
	Zekerheid (certainty):	Gevolgen (danger):	Matrix-waarde:
Groot w. risico	almost never	a great deal	(1)
	almost never	some/not much	(2,5)
	usually/sometimes	a great deal	(2,5)
Matig w. risico	very	a great deal	(4)
	almost never	no danger	(4)
	usually/sometimes	some/not much	(6,25)
Klein w. risico	very	some/not much	(10)
	usually/sometimes	no danger	(10)
	very	no danger	(16)

*) Lees $1 \times 4 = 4$

Op deze manier ontstaat een ordinale schaal.

Ook Cunningham is zich bewust van het feit dat zijn methode nogal arbitrair is; zijn schaal is namelijk gebaseerd op de volgende twee veronderstellingen:

1. dat zowel 'zekerheid' als 'gevolgen' eenzelfde gewicht hebben, en
2. dat de afstand tussen de schaalpunten op beide schalen gelijk is.

* Wij geven de vraagstellingen weer in de oorspronkelijke (meestal Engelse) taal: een vertaling zou o.i. afbreuk kunnen doen aan de zuiverheid van de operationalisaties.

Veel onderzoekers hebben in navolging van Cunningham deze meetmethode gebruikt (zie o.a. Copley en Callom (1971), Hisrich et al. (1972), Strazzieri en Green (1979) ofwel een gelijkende variant, zoals bv. zekerheid en *belang* (Arndt (1967b), Schiffman (1972)).

Hoewel men het al of niet eens kan zijn met de vooronderstellingen die ten grondslag liggen aan de Cunningham-methode: de populariteit van zijn benadering lijkt vooral te danken te zijn aan de inzichtelijkheid van de operationalisatie.

Immers, de methode is toepasbaar op alle risicosoorten (zie 2.3), en leent zich bovendien gemakkelijk voor een schriftelijke of mondelingen vraagstelling in een marktonderzoek.

Het is vooral Bettman (1972) geweest, die een nieuwe impuls gaf aan het conceptualiseren en meten van waargenomen risico. Hij introduceerde de eerder genoemde termen 'inherent' en 'gehanteerd' risico, waarbij hij een uitgebreide, paarsgewijze vergelijking als meetinstrument gebruikt:

'In each case, a brief description of the concept of risk, outlining the ideas of uncertainty, consequences, economic, and social risks, was given. Then, for inherent risk the respondent was asked, for pairs of product types, to rate which product type of the pair would be more risky to shop for in an imaginary store where all packages were labeled only with a product type and size, no brand labels or prices. Respondents were told that this store had a full selection of all the usual brands for each product type; they were just to assume they could not tell which brand was which. After choosing which product type was more risky, respondents then rated how much more risky that type was on a scale ranging from 0 (equally risky) to 9 (very much more risky). Handled risk was measured in essentially the same way, but by evoking a different choice situation. The respondent was asked to rate which product type in a pair was more risky to shop for in her own usual grocery store. Hence the effect of brand knowledge and information is included.' (Op. cit. p. 394.)

Het verschil tussen beide risicocomponenten bestaat dus uit de verschillende *keuzesituaties* (denkbeeldige winkelsituatie versus de gewone winkelsituatie). Naar ons idee levert het meten van inherent risico veel problemen op voor de ondervraagden, omdat het voorstellingsvermogen c.q. abstractieniveau vrij behoorlijk dient te zijn.

Wat te denken van de volgende vraag, bestaande uit 58 (!) woorden, die Bettman hanteerde in zijn onderzoek uit 1975 om inherent risico te meten:

'You feel that: about (5%, 35%, 65%, 95%) of the brands in this product class would be acceptable to you if you were going to use them, *and* it is (not at all, somewhat, moderately, very) important to you that you choose the most satisfactory brand in this product class.
To you, choosing a brand of this product class in the imaginary store would be:

Not risky
at all 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Exceptionally risky'

(Bettman, 1975,p.382.)

Ook Deering en Jacoby (1972) hebben een poging gedaan om waargenomen risico op een nieuwe wijze te meten. Zij lieten 230 huisvrouwen een twintigtal produkten beoordelen aan de hand van een negenpunts schaal bij tien vragen (zie tabel 2).

Aan de hand van deze uitkomsten stelden de onderzoekers drie verschillende (samengestelde) maatstaven voor waargenomen risico samen.

Tabel 2. Operationalisaties van Deering & Jacoby (1972)
Tien vragen voor het meten van waargenomen risico:

A. How certain are you that a brand name of this product you haven't tried will work as well as your present brand?

B. We all know that not all products work as well as others; compared to other products, how much danger would you say there is in trying a brand of this product that you have never used before?

C. How confident would you are about judging the quality of the product?

D. Buying a product that gives you good results may be more important for some products listed than for others. How important would you say it is for this product to satisfy you?

E. The investment you make when you buy a product includes your time and energy as well as money. In terms of the time, money, and overall effort required to buy this product, how much would you say you invest?

F. Can most shoppers guess ahead of time how dependable this product will be if it is used over and over again?

G Before buying this product, can almost anyone tell how good its materials are and how well it's put together?

H. Can almost any shopper predict what the bad results will be if this product fails?

I. In general, does this product tend to fulfil your expectations?

J. Is it obvious why someone like yourself would want this product?

De eerste maatstaf bestaat uit de originele Cunningham-methode, t.w. zekerheid × gevaar (oftewel A × B, zie tabel 2).

In de tweede maatstaf ligt het accent op een combinatie van onzekerheid, prestatie- en financieel risico

$$\left(C \times \frac{D+E}{2}\right);$$

de derde methode is de meest complete meting van risico en bevat gevolgen, onzekerheid en de mate en het type van bevrediging die een koopdoel verschaft:

$$\frac{(D+E)}{2} \times \frac{(F+G+H+I+J)}{5}.$$

Na een vergelijking van deze drie metingen vonden de onderzoekers correlatiecoëfficiënten van .87 en hoger (!) tussen de methoden onderling.

Vervolgens werd het gemiddelde berekend van de drie metingen voor elk van de twintig produkten, waarna een rangordening ontstond van produkten met een groot waargenomen risico naar produkten met een klein risico (zie 4.2, voor de risico's per produktklasse).

De hoge correlatiecoëfficiënten duiden erop dat de drie gehanteerde meetmethoden globaal hetzelfde theoretische concept meten. Wat echter ontbreekt in dit verband is (wederom) een goede theoretische definitie van waargenomen risico: men gaat te snel naar de vraagstellingen toe.

Een ander discussiepunt betreft de vraag of de risicocomponenten nu opgeteld of vermenigvuldigd zouden moeten worden, m.a.w. is het 'onzekerheid *maal* gevolgen' ofwel 'onzekerheid *plus* gevolgen'? Eenzelfde vraag doet zich voor bij het onderzoek naar 'decisionmaking'. Bettman (1973) concludeert aan de hand van een uitvoerige multivariatie-analyse dat beide benaderingen gebruikt kunnen worden, dus zowel de multiplicatieve als de additieve.

Wanneer we nu de 'state of the art' overzien, dan is het duidelijk dat er nog wel het een en ander rammelt in de operationalisaties van waargenomen risico.

Ross (1974) noemt een aantal zinvolle tips met betrekking tot het meten van risicoperceptie in de toekomst, die wij niet onvermeld willen laten. Hij adviseert om *onopvallende* of *indirecte* meetmethoden te hanteren, omdat de gebruikelijke schalen de ondervraagden sensitief kunnen maken. Immers, deze directe metingen van risico kunnen de perceptie van de ondervraagden dusdanig beïnvloeden, dat ze zich inderdaad in een riskante situatie denken te bevinden, terwijl ze zonder deze directe meting de situatie wellicht niet als riskant zouden waarnemen.

Een tweede punt betreft het tijdstip van meting van risicoperceptie en die van risicobehandeling: het liefst dient men deze concepten *gescheiden* van elkaar te meten, omdat anders het verband tussen beide begrippen door de respondent al te gemakkelijk zal worden herkend, met als gevolg een sterke rationalisatie. Ten slotte adviseert hij om vaker gebruik te maken van *experimentele manipulatie* van risico, waardoor duidelijker conclusies uit de resultaten getrokken kunnen worden.

In de praktijk is het echter niet altijd mogelijk om gebruik te maken van een laboratoriumtest, vooral als het grote aantallen ondervraagden betreft.

Wij willen hierbij aantekenen dat het meten van risicoperceptie en risicobehandeling in feite op ten minste twee manieren zou moeten plaatsvinden, nl. op het *cognitieve* niveau (= vragenlijst) èn op het *gedragsmatige* niveau (= gedragsobservatie/-registratie).

4. DE INVLOED VAN PERSOONLIJKHEIDSTREKKEN EN SITUATIONELE ELEMENTEN OP HET WAARGENOMEN RISICO

4.1. *Cognitieve stijl, categoriebreedte en zelfvertrouwen*

Bij de bestudering van het consumentengedrag hebben sommige onderzoekers gepoogd om waargenomen risico te verklaren door middel van een of meer 'trekken'. De meest voorkomende begrippen in dit verband zijn: cognitieve stijl, categoriebreedte en zelfvertrouwen. Witkin (1964) geeft de volgende algemene definitie van *cognitieve stijl:* 'Cognitive styles are the characteristic, self consistent ways of functioning shown by the person in the cognitive sphere', m.a.w. de specifieke manier waarop iemand denkt, redeneert, gelooft en zijn kennis gebruikt.

Cox en Rich (1964) vonden dat sommige mensen een duidelijk consistent gedrag vertoonden m.b.t. risicoreductie: elke groep consumenten had een specifieke (= cognitieve) stijl om de onzekerheid te verminderen, zoals bv. informatievergaring of het steunen op eigen ervaring uit het verleden (zie ook Pin-

son (1975)). Een begrip dat verband houdt met het voorgaande is *categoriebreedte,* geïntroduceerd door Pettigrew (1956). Popielarz (1967) onderzocht dit concept m.b.t. risiconame door consumenten, en vond dat mannen die de neiging hebben om produktverschillen over het hoofd te zien (zg. 'broad categorizers') eerder genegen zijn om nieuwe produkten te proberen. Bij vrouwen vond hij echter een omgekeerde relatie: 'broad categorizers' waren juist minder genegen om iets nieuws te proberen.

Cox (1967a) merkte een verschil in cognitieve stijl op bij zijn twee onderzochte (huis)vrouwen: de een zocht nogal eens informatie om opheldering te krijgen ('clarifier') voor de risicoreductie, terwijl de ander juist probeerde om mogelijk verwarrende informatie van buitenaf tegen te houden ('simplifier').

Het begrip *zelfvertrouwen* (Eng. 'self-confidence, self-esteem') ten slotte, heeft ook nogal wat aandacht gekregen m.b.t. risico. Een aardige omschrijving van dit begrip is de volgende:

'The evaluation which the individual makes and customarily maintains with regard to himself: it expresses an attitude of approval or disapproval, and indicates the extent to which the individual believes himself to be capable, significant, successful and worthy' (Coopersmith, 1967).

Een aantal schrijvers maken een onderscheid tussen algemeen en specifiek zelfvertrouwen, zoals bv. Hisrich et al. (1972).

Het algemene zelfvertrouwen kan gezien worden als een persoonlijkheidstrek, zoals hierboven omschreven door Coopersmith. Specifiek zelfvertrouwen daarentegen is een situationeel bepaalde factor, zoals bv. het zelfvertrouwen dat men heeft om een goede winkel te kiezen voor de aankoop van een produkt. Dit komt dicht in de buurt van de eerder besproken risicocomponent '(on)zekerheid' (zie 2.1.).

Uit het onderzoek van Hisrich et al. (op. cit.) blijkt dat het verband tussen specifiek zelfvertrouwen en risicoperceptie sterker is dan tussen algemeen zelfvertrouwen en risicoperceptie. Dit ligt in de lijn van de verwachting, doch ook hier hangt alles sterk af van de meetmethoden die zijn gebruikt.

Schweiger (1977) concludeert dat het specifieke zelfvertrouwen een directe invloed heeft op het waargenomen risico.

Wij zijn echter geneigd te constateren dat de begrippen zelfvertrouwen en (on)zekerheid zéér dicht bij elkaar liggen, zodat een onderscheid t.a.v. risicoperceptie overbodig is.

Mathews et al. (1971) hebben specifiek de relatie onderzocht tussen waargenomen risico en enige persoonlijkheidstrekken. Ze vonden dat consumenten die bereid waren tot het nemen van grote risico's t.a.v. de merkkeuze, *impulsiever* waren dan anderen. Aldus was het mogelijk een onderscheid te maken tussen risiconemers en risicomijders: risicomijders bv. bleken in mindere mate bereid tot het kopen van luxe goederen op afbetaling dan risiconemers. Ook bleek dat in de hoge leeftijdsgroepen het relatieve aantal risicomijders toenam.

We zien dat de invloed van persoonlijkheidskenmerken op waargenomen risico nogal onduidelijk uit de literatuur naar voren komt. Zo er al verbanden bestaan dan zijn die nogal vaag, indirect en niet steeds met elkaar in overeenstemming.

4.2. *De invloed van de produktklasse en de koopsituatie*

Wanneer we waargenomen risico gaan bezien als een situationele en/of een produktgebonden variabele, dan blijken de verbanden hechter en consistenter te zijn.

Taylor (1974) en Vincent en Zikmund (1979) benadrukken de keuze- of koopsituatie als de belangrijkste determinant van het waargenomen risico.

Dit betekent dat empirisch onderzoek naar situatie-specifiek waargenomen risico van groot belang kan zijn voor diegenen, die belangstelling hebben voor een specifieke produktklasse (bv. marketing managers). Er is dan ook nogal wat onderzoek verricht op dit terrein.

We konden echter geen structurele analyse vinden m.b.t. waargenomen risico over de verschillende produktklassen heen: zijn er wellicht produktklassen die meer risico met zich meebrengen dan andere? Zo ja, wat is dan de onderliggende structuur of taxonomie?

Het schijnt dat de meeste onderzoekers een of meer produkten hebben genomen op basis van intuïtieve, gevoelsmatige criteria: het is immers duidelijk dat het kopen van een auto waarschijnlijk meer risico met zich meebrengt dan het kopen van een pond suiker. In tabel 3 geven wij een taxonomie van onderzochte produktvelden en hun waargenomen risico, samengesteld op grond van acht verschillende onderzoeken (de belangrijkste).

Een waarschuwing bij de interpretatie van de tabel is wel op haar plaats, omdat steeds verschillende meetmethoden zijn gehanteerd. Wij hebben bovendien rangorde-maatstaven en scores getransformeerd in drie relatieve risicointensiteiten: hoog, gemiddeld en laag. Ruwweg ontstonden er tien produktvelden: automobielen, cosmetica, (levens)verzekeringen etc., gezondheidsprodukten, produkten in en om het huis, recreatie binnen- en buitenshuis, voedings- en genotmiddelen, kleding en schoeisel, reinigers en overige. De volgorde in onze taxonomie is niet geheel toevallig: automobielen en cosmetica hebben, zeker bij vrouwen, in het algemeen een groter waargenomen risico dan bv. een balpen. Het zal evenwel sterk van de doelgroep afhangen welke produkten men als riskanter ervaart dan andere: we dienen ons te realiseren dat de uitkomsten in tabel 3 vooral bepaald zijn door de antwoorden van studenten en (huis)vrouwen die bereid waren mee te doen met het onderzoek!

Hier en daar is duidelijk dat de uitkomsten elkaar tegenspreken: bijvoorbeeld bij hoofdpijnstillers en levensverzekeringen. Dit kan mede veroorzaakt zijn doordat produkten soms in hun onderlinge afhankelijkheid zijn beoordeeld op waargenomen risico: het ontbreken resp. vóórkomen van een of meer produkten heeft dan een sterke invloed op de uitkomsten van het onderzoek.

Een ander punt is, dat niet alle onderzoekers de verschillende risico*soorten* apart hebben gemeten; het blijft daarom de vraag in hoeverre er sprake is van risico's van geld/tijd/moeite, object- of subjectrisico's bij elk der produktklassen.

Wat echter wèl duidelijk naar voren komt, met name bij Jacoby en Kaplan (1972), is dat de *waarde* van het produkt in het algemeen samenhangt met de hoeveelheid risicoperceptie: hoe duurder, des te meer risico ervaart men. Dit is ook duidelijk terug te vinden in de grote hoeveelheid duurzame consumptiegoederen in de linker kolom ('hoog') van tabel 3.

Deze rangordening-naar-waarde-hypothese wordt ook bevestigd door Perry en Hamm (1969), alsmede door Hisrich et al. (1972, p. 437).

De relatie 'Duur'→'meer risico' is echter eenzijdig: 'Goedkoop' betekent niet automatisch 'minder risico'.

Het is duidelijk dat produktvelden als cosmetica en gezondheidsprodukten meestal aanzienlijk goedkoper zijn dan duurzame consumptiegoederen; toch scoren deze hoog op risico. Toekomstig onderzoek zou zich moeten richten op het nauwkeuriger in kaar brengen van de verschillende risicosoorten die behoren bij de desbetreffende produktvelden.

Situationele factoren kunnen veel invloed uitoefenen op de wijze waarop men risico's waarneemt.

Cox en Rich (1964), Spence et al. (1970), Hisrich et al. (1972) en Vincent en Zikmund (1979) hebben aangetoond dat ook andere (omgevings)factoren naast het produkt zelf invloed hebben op de risicoperceptie, t.w. de *plaats en wijze van winkelen.*

Cox en Rich veronderstelden dat winkelen via de telefoon risico oproept en dat daarom sommige vrouwen ervan worden weerhouden om op deze wijze te 'winkelen'. Ook Spence et al. vonden dat kopen via postorders riskanter werd ervaren dan via een winkel of een verkoper.

Hisrich et al. bemerkten dat wanneer men onbekend is met het produkt zelf door bv. een lage aankoopfrekwentie (vloerbedekking), men afgaat op de *winkel* als een soort 'surrogaat'-risicoreductor.

Het onderscheid van Bettman (1972) tussen inherent en gehanteerd risico haakt in feite ook in op de invloed van de aankoopsituatie (denkbeeldige koopsituatie versus gewoonlijke winkel).

Een derde en wellicht zeer belangrijke factor is de *interpersoonlijke situationele omstandigheid* van de aankoop.

Vincent en Zikmund (1979) ontdekten dat het sociale risico van het kopen van een electrisch mes voor eigen gebruik veel lager was dan als huwelijkscadeau voor een vriend.

Bij het financiële risico bleek dat deze relatie juist omgekeerd was: de koopprijs van een produkt voor eigen gebruik bracht een hoger risico teweeg dan in geval van een geschenk voor een ander. De beide andere risicosoorten (prestatie- en fysiek risico) toonden geen grote verschillen in beide situaties.

Een ander psychologisch verschijnsel, nl. *risicoverschuiving* (Eng. 'risky shift'), kan ook invloed hebben op de risicoperceptie. Dit begrip stamt uit de sociale psychologie; in het kort komt het hierop neer, dat het individu geneigd zal zijn meer risico te nemen in een groep mensen dan wanneer hij alleen is.

Kogan en Wallach (1964) verklaarden de grotere risicogeneigdheid van groepen door aan te nemen dat in groepen de verantwoordelijkheid voor beslissingen diffuus wordt, m.a.w. de individuele verantwoordelijkheid wordt minder.

Woodside (1972, 1974) paste deze theorie toe op het consumentengedrag. Een aantal consumenten werd een hypothetische koopsituatie voorgelegd, waarbij een keuzedilemma werd gepresenteerd.

De keuzedilemma's (auto's, stoomstrijkijzers en oploskoffie) waren speciaal gemaakt en wel zodanig, dat de keuze van het riskantere alternatief een po-

Tabel 3. Indeling van produkten naar hoeveelheid waargenomen risico

	Scores/waarden van waargenomen risico (gemiddelde waarden, zie voetnoot)		
Produktklassen	hoog	gemiddeld	laag
Auto's	auto's (4,6,7,8) autobanden (4)	accu's (7) benzine (4)	
Cosmetische produkten	haarlak (4) Eau de cologne (7)	haarlak (7) naar de kapper gaan (7) deodorant (8) contactlenzen (7) scheermesjes (8)	
Verzekering, etc.	levensverzekering (3,8) beleggingsfonds (3)	ziekenhuisverzekering (3)	levensverzekering (7)
Gezondheidsprodukten	hoofdpijnstillers (1,3,4) tandpasta (5) vitamines (3)	hoofdpijnstillers (5)	hoofdpijnstillers (8) tandpasta (8) vitamines (8)
Produkten rondom het huis	appartement (7) wasautomaat (6) koelkast (4) electr. naaimachine (3) microgolf-ovens (3) barkrukken (2) aluminium kozijnen (3)	el. mixers (6) wandbekleding (6) tafellamp (4) kaarttafel met stoelen (2) metalen tuinmeubels (2) electr. grasmaaier (3)	stoomstrijkijzer (6) sauspan (4) metalen tuinmeubels (3)
Recreatie binnen- en buitenshuis	kleuren-TV (7,8) stereo HIFI set (7) HIFI elpee (3)	camera (6) kleuren-TV (2) kleinbeeld-TV (3) grammofoon (4) Monopoly-spel (3)	golfclubs (7) Italiaanse maaltijd (7) tijdschrift (7) spel kaarten (8)
Voedings- en genotsmiddelen	bier (7) oploskoffie (5) margarine (5)	bier (4,5) sigaretten (7) Bourbon-whiskey (3)	consumptie-ijs (4) zout (2,4) pak spaghetti (1,4,5) koffie (6) verse aardbeien (3)
Kleding, schoeisel, etc.	pak (7,8) bontjas (6) winterjas (7)	zwarte sokken (7) nette schoenen (4,8) ondergoed (4) sportief jasje (8)	T-shirt (7) zonnebril (7) polshorloge (7) regenjas (4) kinderschoenen (3)
Reinigers, etc.	wasmiddel (4) vloerwrijfwas (4)	wasverzachter (1,4) meubelwas (5) schoenpoets (7)	wasverzachter (5) papieren handdoeken (5)
Overige		tulpenbollen (3) wegwerpluiers (3)	schrijfpapier (2) kerstkaarten (3) balpen (7) badhanddoeken (4) tweepersoonslakens (3)

BRONNEN:
(1) = Cunningham (1967), tabel 3.2.; gebaseerd op de kolommen 'hoog' en 'laag'.
(2) = Zikmund & Scott (1973), tabel 1; gemiddelde alg. risicoscores: 1,74-1,56 = laag/2.42-2.95 = gem./3.00-5.69 = hoog.
(3) = Spence et al. (1970), tabel 2; winkel/verk. risico rangord.: 1-2.66 = laag/2.67-3.22 = gem./3.23 + = hoog.
(4) = Deering & Jacoby (1972), tabel 2; rangorde: 1-6 = laag, 7-13 = gem./14-20 = hoog (gem. risicowaarden)
(5) = Bettman (1972), tabel 1; inherent risico (gem.) (gemidd. waarden): −2.186 - −.919 = laag/ −.594-.716 = gemid./ 1.103-1.706 = hoog.
(6) = Woodside (1974), tabel 2; rangorde v. risicowaarden: 7,8 = laag/4-6 = gemid./1-3 = hoog
(7) = Perry & Hamm (1969), tabel 1; socioec. risico rangorde: 17-25 = laag/9-16 = gem./1-8 = hoog
(8) Jacoby & Kaplan (1972), tabel 3; rangorde: 9-12 = laag/5-8 = gem./1-4 = hoog.

N.B. Onderzoeknrs. 1,2,4,5 en 6 werden uitgevoerd bij vrouwen; de nrs. 7 en 8 onder mannen, terwijl nr. 3 bij zowel mannen als vrouwen werd gehouden.

tentieel grotere beloning/bevrediging zou opleveren, en vice versa.

Woodside mat vervolgens het verschil tussen de individuele scores op de risicoperceptie vóór en ná een groepsdiscussie, en concludeerde dat er een toename was van de bereidheid om riskantere produkten te kiezen, m.a.w. de hoeveelheid waargenomen risico was minder geworden door de discussie. We kunnen veronderstellen dat de discussie opheldering heeft gegeven ('clarifier'-funktie, zie 4.1.). Oudere vrouwen evenwel hadden na de discussie een hoger waargenomen risico dan ervoor, vergeleken met jongere vrouwen ('simplifier'-functie).

Voorafgaand aan het experiment had Woodside de hoeveelheid risicoperceptie van een aantal produktvelden d.m.v. rangordening laten vaststellen. Het bleek dat drie van de vier risicoverschuivingen optraden bij produkten met een *gemiddelde* hoeveelheid risico, t.w. stoomstrijkijzers, camera's en mixers, en dus niet, zoals was verwacht, bij produkten met een groot waargenomen risico.

Met andere woorden: consumenten zijn wellicht eerder terughoudender dan dat zij risico's nemen na een groepsdiscussie over produkten met een groot waargenomen risico (zie ook Reingen (1974).

Wanneer we denken aan toepassingen van risicoverschuiving in de marketingpraktijk, dan zien we dat verkoopmaatschappijen die gebruik maken van zg. house-parties, hierop inspelen.

Het zal echter steeds afhangen van de produktklasse òf en in welke mate er risicoverschuiving zal optreden.

Resumerend kunnen we vaststellen dat de aankoopsituatie een belangrijke rol kan spelen bij de bepaling van het waargenomen risico. Het probleem is echter, hoe we moeten komen tot een systematisch inzicht in deze situationele variabelen.

Belk (1975) noemt de volgende vijf groepen van situationele kenmerken die een rol kunnen spelen:

1. *Fysieke omgeving* (locatie, achtergrond, weer, materiaal, etc.).
2. *Sociale omgeving* (aanwezigheid van anderen, interacties, etc.).
3. *Tijdsperspectief* (koopfrequentie, gebruiksfrequentie, tijdsdruk).

4. *Taakomschrijving* (winkelkeuze, eigen gebruik of cadeau, etc.).
5. *Antecedenten* (stemming, vermoeidheid, ziekte, etc. v.d. koper).

Slechts enkele van deze kenmerken vinden we terug in de literatuur rondom risicoperceptie, nl. 2) en 4). Het is te hopen dat de andere kenmerken in de toekomst ook aan bod zullen komen bij onderzoek naar de invloed van de situatie op risicoperceptie.

5. RISICOBEHANDELING: RISICOREDUCTIE EN RISICOVERGROTING

Meestal bestaat risicobehandeling uit een proces van risicovermindering vóór de aankoop, doch soms kan er ook sprake zijn van risicovergroting; hierop komen we terug in 5.2.

Meestal betekent risicobehandeling het vergroten van de zekerheid of het verminderen van de koopaspiratie ('buying goals', zie Cox 1967a)). Men verkiest meestal het vergroten van de zekerheid, en zodoende wordt volgens Cox risicobehandeling gelijk aan *informatiebehandeling:* 'Informatie is het tegengif voor onzekerheid' (op. cit. p. 81).

Wanneer we risicobehandeling als informatiezoekgedrag beschouwen, zien we dat Berlyne's theorie van de 'nieuwsgierigheid' (1960) een aantal onderzoekers heeft geïnspireerd op het terrein van de risicoperceptie. Deze theorie, die wat verfijnd is door Howard en Sheth (1969), veronderstelt een U-vormig verband tussen 'stimulus ambiguïteit' (=hoeveelheid risico) en 'arousal' (= mate van prikkeling die aanzet tot informatievergaring). In figuur 3 zien we het veronderstelde verband grafisch weergegeven.

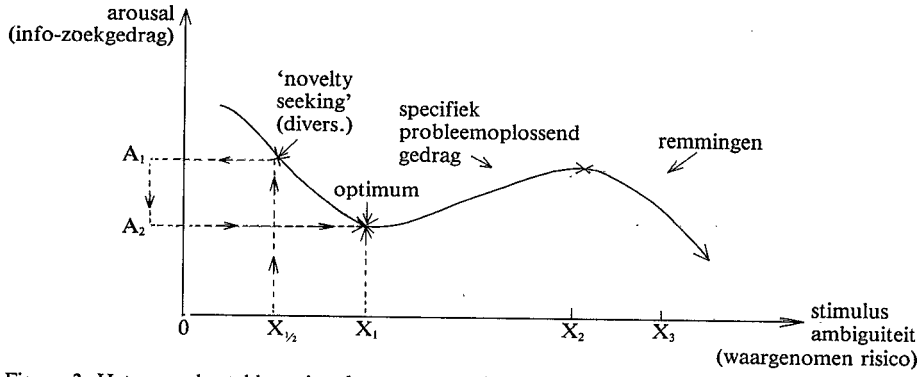
Volgens de theorie zal het individu steeds proberen het optimumniveau te bereiken, alwaar het arousal het laagste is (A_2,X_1). Het gebied tussen O en X_1 in de figuur kan omschreven worden als 'novelty seeking' (zoeken naar nieuwe prikkels): de hoeveelheid waargenomen risico is zeer laag (stel, $X_{1/2}$) en verveling t.a.v. een produkt of merk kan optreden. In dat geval kan zich zg. diversief zoekgedrag voordoen, omdat het arousalniveau (te) hoog is, bv. A_1. De consument zal doorgaan tot het optimum A_2 wederom bereikt is. In dit geval is er sprake van *risicovergroting.*

Tussen X_1 en X_2 wordt in toenemende mate risico waargenomen en dus zal het info-zoekgedrag toenemen, gericht op het oplossen van het specifieke probleem (*risicoreductie*).

Voorbij het punt X_2 zal, volgens de theorie, de hoeveelheid waargenomen risico zó groot worden, dat remmingen en/of defensief gedrag zal optreden. Dit kan zich voordoen, wanneer de koopsituatie te complex is om te hanteren: denk bijvoorbeeld aan het afzeggen van een afspraak met een makelaar om een te koop staand huis te bezichtigen, wanneer men bang is voor een té impulsieve aankoop.

De hierboven beschreven theorie van Berlyne verschaft een basis voor de bestudering van risicobehandeling.

Sommige onderzoeken tonen echter aan, dat informatiebehandeling weliswaar een belangrijke, doch niet de enige manier is om risico's te verkleinen, zoals we hierna zullen zien.



Figuur 3. Het veronderstelde verband tussen arousal en stimulus ambiguïteit*)

5.1. *Risicoreductie en risicoreductoren*

Er zijn een aantal manieren om risico's te behandelen.

Roselius (1971) noemt elf reductoren; de meest uitgebreide lijst die we in de literatuur konden vinden. Opmerkelijk is wel, dat de *reclame* als informatiebron door hem niet werd genoemd (!).

Gebruikmakend van de informatiebronnen-typologie van Andreasen (1968) en de Roselius-lijst hebben wij een lijst van risicoreductoren samengesteld (zie tabel 4).

Tabel 4. Typologie van risicoreductoren

1. Onpersoonlijk/partijdig (goede image van bekend merk, advertenties).
2. Onpersoonlijk/onafhankelijk (Consumentengids, etc.)
3. Persoonlijk/partijdig (image van buurtwinkel, verkoopbedienden).
4. Persoonlijk/onafhankelijk (mond-tot-oor reclame van bekenden).
5. Directe observatie/eigen ervaring (merkentrouw, winkelen, gratis monster, geld-terug-garantie, duur = goed, etc.).

Men kan zowel volgtijdelijk, gelijktijdig of afzonderlijk gebruik maken van deze reductoren. Uit Roselius' onderzoek blijkt dat merkentrouw, eigen ervaring en het kiezen van een bekend merk de meest populaire risicoreductoren zijn; duur = goed was het minst geliefd. Ook Bauer (1960) constateerde reeds dat merkentrouw een goede manier is om risico's van aankopen te verminderen (p. 25).

Cunningham (1967c) alsmede Arndt (1967b) constateerden dat personen met een groot waargenomen risico t.a.v. oploskoffie een grotere merkentrouw vertoonden en dus ook in mindere mate een nieuw merk zouden accepteren.

Het gebruik van een bepaalde reductor hangt ook af van de *hoeveelheid* waargenomen risico die men ervaart: Lutz en Reilly (1973) vonden dat, inge-

*) De vorm van de afgebeelde curve zal verschillen per individu alsmede per situatie. Voor een soortgelijk verband tussen waargenomen conflict en complexiteit van de keuzesituatie, zie Hansen (1972), p. 160.

val het prestatierisico als gering of gemiddeld werd ervaren, men het produkt eenvoudigweg kocht (probeersel); doch wanneer het prestatierisico groot werd geacht, de meesten 'directe observatie' en 'eigen ervaring' prefereerden.

We veronderstellen dat 'aankoop-probeersels' alleen zullen optreden wanneer de aankoopprijs relatief (d.w.z. vanuit de koper gezien) laag is. Wanneer men de prijs ziet als gemiddeld of hoog, zullen er andere reductoren gebruikt worden.

In het voornoemde onderzoek van Lutz en Reilly bleek, dat het tijdschrift 'Consumer Reports' (met de Consumentengids vergelijkbare uitgave in de USA) alleen geraadpleegd werd in geval van een groot prestatie-risico (en, naar ons idee, in geval van produkten met een hoge aanschafwaarde).

Het verkrijgen van *nadere informatie* omtrent produkten is, zoals we reeds eerder zagen, een van de meest geliefde reductoren. Wanneer er geen eigen ervaring met een produkt bestaat en directe observatie niet mogelijk is, zal men meestal hierop zijn aangewezen (zie 1 t/m 4 in tabel 4). Het blijkt nu in 't algemeen dat de (informele) *mond-tot-oor communicatie* meestal de belangrijkste/doorslaggevende informatie bevat, speciaal in hoog-riskante aankoopsituaties.

Veel auteurs vonden bijvoorbeeld ondersteuning voor de veronderstelling dat, hoe groter het waargenomen risico in een bepaalde aankoopsituatie is, des te groter ook het belang van persoonlijke beïnvloeding wordt (zie bv. Perry en Hamm (1969), Roselius (1971), Lutz en Reilly (1973), Sheth en Venkatesan (1968) en Arndt (1967b).

5.2. *Risicovergroting*

Gestimuleerd door de eerder besproken theorie van Berlyne, hebben een aantal onderzoekers het idee geopperd dat consumenten in sommige situaties de hoeveelheid risico bewust willen *vergroten* en zelfs soms dissonante informatie opzoeken. Dit zal alleen gebeuren in een situatie van verveling, bij routineaankopen: men gaat doen aan zg. 'diversieve exploratie'.

Met betrekking tot waargenomen risico zijn de volgende twee publicaties van belang:

Copley en Callom (1971) hebben als eerste dit verschijnsel onderzocht in hun studie naar het gedrag van industriële inkopers. Ze veronderstelden dat ter linkerzijde van X_1 (zie figuur 3) de ambiguïteit (= waargenomen risico) zo gering is, dat de inkoper zijn zoekgedrag zal opvoeren als gevolg van zijn nieuwsgierigheid en zijn behoefte om de sleur te doorbreken van steeds zaken doen met dezelfde leveranciers. Inderdaad vonden zij een groep die zich gedroeg als de curve van Berlyne, doch het was in totaal slechts 8% die zich aldus manifesteerde...

Deering en Jacoby (1972) vonden bevestiging voor hun hypothese dat consumenten in het algemeen de grootste voorkeur vertonen voor produkten met een *gemiddeld* risiconiveau (zie fig. 3 bij X_1, het optimum). Zowel produkten met een gering als produkten met een groot risiconiveau werden vermeden.

Risicovergroting zal, indien het al voorkomt, alleen plaatsvinden in situaties van verveling met hetzelfde produkt/aankoopkanaal; dit kan zijn in geval van

routinematige aankopen van dagelijkse produkten (boodschappen).

6. TOEKOMSTPERSPECTIEF

In dit literatuuroverzicht hebben wij een zo breed mogelijk overzicht gepresenteerd van het begrip waargenomen risico, waarbij hier en daar toepassingen in het marktonderzoek werden gerefereerd.

Momenteel is het wat stil geworden in de vakliteratuur rondom het onderwerp waargenomen risico. Zo het nog aan de orde komt, wordt meestal de Cunningham-methode gebruikt, zonder verdere theoretische discussie (zie bv. Upah (1979)).

De talrijke conceptuele en methodologische haken en ogen die kleven aan elk onderzoek naar dit verschijnsel zijn daar waarschijnlijk debet aan.

Wil risicoperceptie een alom gehanteerd begrip gaan worden, dan dienen er wel oplossingen gevonden te worden voor de hierboven geschetste fundamentele problemen: pas dan kan verder worden gegaan met inhoudelijk onderzoek op het gebied van de verschillende produktvelden.

BIBLIOGRAFIE

Andreasen, A., 1968. Attitudes and customer behavior: a decision model. In: Kassarjian & Robertson (eds.) *Perspectives in Consumer Behavior.* Scott, Foresman & Comp., 498-510.

Arndt, J., 1967a. Word of mouth advertising and informal communication. In: Cox (ed.) *Risk taking and information handling in consumer behavior,* Boston. 188-239.

Arndt, J., 1967b. Perceived risk, sociometric integration, and word of mouth in the adoption of a new food product. In: Cox (ed.) *Risk taking and information handling in consumer behavior,* Boston, 289-316.

Bauer, R.A. 1960. Consumer behavior as risk taking. In: *Proceedings of the 43rd Conference of the A.M.A.,* 389-398.

Belk, R.W., 1975. Situational variables and consumer behavior. *Journal of Consumer Research II,* 157-167.

Berlyne, D.E., 1960. Conflict, arousal and curiosity. New York, McGraw Hill.

Bettman, J.R., 1972. Perceived risk: a measurement methodology and preliminary findings. In: *Proceedings of the association of consumer research,* 394-403.

Bettman, J.R., 1973. Perceived risk and its components: a model and empirical test. *Journal of Marketing Research X,* 184-190.

Bettman, J.R., 1975. Information integration in Consumer risk perception: a comparison of the two models of component conceptualisation. *Journal of Applied Psychology 60-III,* 381-390.

Coopersmith, S., 1967. The antecedents of self-esteem. San Francisco W.H. Freeman & Co.

Copley, T.P. en Callom, F.L., 1971. Industrial search behavior and perceived risk. In: *Proceedings of the Association of Consumer Research,* 208-231.

Cox, D.F., 1967. Risk taking and information handling in consumer behavior. Boston, Harvard University.

Cox, D.F., 1967a. Risk handling in consumer behavior: an intensive study of two cases. In: Cox (ed.): *Risk taking and information handling in consumer behavior.* Boston, Harvard University.

Cox, D.F., 1967b. Risk taking and information handling in consumer behavior. In: Cox (ed.): *Risk taking and information handling in consumer behavior,* 604-639.

Cox, D.F. en Rich, S.U., 1964. Perceived risk and consumer decision making: the case of telephone shopping. *Journal of Marketing Research I,* 32-39.

Cunningham, S.M., 1967a. The major dimensions of perceived risk. In: Cox (ed.): *Risk taking and information handling in consumer behavior,* Boston, Harvard University.

Cunningham, S.M., 1967c. Perceived risk and brand loyalty. In: Cox (ed.): *Risk taking and information handling in consumer behavior,* Boston, Harvard University.

Deering, B.J. en Jacoby, J., 1972. Risk enhancement and risk reduction as strategies for handling

perceived risk. In: Proceedings of the association of consumer research, 404-416.
Hansen, F., 1972. Consumer choice behavior: a cognitive theory. New York, Free press.
Hisrich, R.D., Dornoff, R.J. en Kernan, J.B., 1972. Perceived risk in store selection. *Journal of marketing research IX,* 435-9.
Howard, J.A. en Sheth, J.N., 1969. The theory of buyer behavior. New York, John Wiley & Sons.
Jacoby, J. en Kaplan, L.B., 1972. The components of perceived risk. In: *Proceedings of the 3rd annual conference of the Association of consumer research,* 382-93.
Kogan, N. en Wallach, M.A., 1964. Risk-taking: a study in cognition and personality. New York, Holt, Rinehart & Winston.
Kogan, N. en Wallach, M.A., 1967. Risk taking as a function of the situation, the person and the group. In: M. Newcomb (ed.): *New directions in psychology III* New York, Holt, Rinehart & Winston.
Lutz, R.J. and Reilly, P.J., 1973. An exploration of the effects of perceived social and performance risk on consumer information acquisition. In: Ward and Wright (eds.): *Advances in consumer Research,* Urbana, 393-405.
Mathews, H.L., Slocum, J.W. en Woodside, A.G., 1971. Perceived risk, individual differences and shopping orientations. In: Gardner (ed.): *Proceedings of the 2nd annual conference of the association of consumer research.*
Perry, M. en Hamm, B.C., 1969. Canonical analysis of relations between socio-economic risk and personal influence in purchase decisions. *Journal of marketing research VI,* 351-4.
Peter, J.P. en Ryan, M.J., 1976. An investigation of perceived risk at the brand level. *Journal of marketing research,* 184-190.
Peter, J.P. en Tarpey, L.X., 1975. A comparative analysis of three consumer decision strategies. *Journal of Consumer Research II,* 29-37.
Pettigrew, T.F., 1956. The measurement and correlates of category width as a cognitive variable. *Journal of Personality* 26, 532-44.
Pinson, C.R.A., 1975. Consumer cognitive styles: review and implications for marketeers. Fontainebleau, Insead Research Papers Series no. 160, 1-32.
Popielarz, D.T., 1967. An exploration of perceived risk and willingness to try new products. *Journal of marketing research IV,* 368-73.
Reingen, P.H., 1974. Comment on Woodside. *Journal of marketing research* XI, 223-4.
Roselius, T., 1971. Consumer rankings of risk reduction methods. *Journal of marketing* 35, 56-61.
Ross, I., 1974. Perceived risk and consumer behavior: a critical review. In: Schlinger (ed.): *Advances in consumer research II,* Urbana Ill., 1-19.
Schiffman, L.G., 1972. Perceived risk in new product trial by elderly consumers. *Journal of marketing research IX, 106-8.*
Schweiger, G., 1977. Perveived risk: a contribution to the interpretation of the savings behavior of private investors. Oslo, *ESOMAR Annual Congress papers II,* 319-339.
Seth, J.N. en Venkatesan, M., 1968. Risk-reduction processes in repetitive consumer behavior. *Journal of marketing research V,* 307-10.
Spence, H.E., Engel, J.F. en Blackwell, R.D., 1970. Perceived risk in mail-order and retail store buying. *Journal of marketing research VII,* 364-9.
Stallen, P.J.M., 1979. De individuele beoordeling van risico's: Een inleidende beschouwing. *TNO Project* VII-11, 411-16.
Stem, D.E., Lamb, C.W. en MacLachlan, D.L., 1977. Perceived risk: a synthesis. *European journal of marketing* 11-4, 312-19.
Strazzieri, A. en Green, R.T., 1979. An exploratory study of consumer behavior patterns in Subsahara Africa. In: *Proceedings of the 8th Annual meeting of the European Academy of Advanced Research in Marketing,* Groningen, C75-98.
Taylor, J.W., 1974. The role of risk in consumer behavior. *Journal of marketing* 38, 54-60.
Upah, G.D., 1979. Applying the concept of perceived risk to buying influence in industrial firms. In: *Advances in consumer research VII* Urbana, Ill., 381-86.
Vincent, M. en Zikmund, W.G., 1979. An experimental investigation of situational effects on risk perception. In: Olson (ed.) *Advances in Consumer research VII,* 125-9.
Vlek, C. en Stallen, P.J.M., 1980. Rational and personal aspects of risk. *Acta Psychologica* 45, 273-300.

12. Het onderkennen van risico in het kopen van goederen als persoonskenmerk

G.J. RANDOE

SAMENVATTING

Het blijft altijd een waagstuk om, als publikaties uit het buitenland zeer omvangrijke voorstudiën bevatten, een toetsingsonderzoek op te zetten op basis van concepten en operationalisaties van die voorstudie.

Dit verslag is het relaas van zo'n waagstuk, dat mislukte. De auteur doet een poging om de mislukking te evalueren. Dank zij een inhoudelijke verkenning van het concept achteraf is het mogelijk enkele waarschijnlijke oorzaken voor de mislukking te achterhalen.

1. INLEIDING

Deze bijdrage bevat een verslag van een programma van toetsing van hypothesen, dat in de periode van 1974 tot 1976 werd uitgevoerd.

Na een omvangrijk onderzoek bleek dat de oorspronkelijke hypothesen geheel opnieuw moesten worden geformuleerd. Een kater dus voor het desbetreffende onderzoekteam. Het is geheel in deze stijl verslagen: het afsluitende hoofdstuk bevat dertien conclusies.

Aanleiding tot het onderzoek vormde een stroming in marketingresearch in de 70-er jaren, die met name in de Verenigde Staten tot een stroom van publikaties heeft geleid.

Het onderzoek was gericht op studie van het oriëntatiegedrag bij het doen van 'grote' gezinsaankopen.

Deze studies gaan terug tot het bekende onderzoek van Kats en Lazarsfeld (twop-step-flow of communications-theorie, zie (8)).

Donald Cox stelde in 1967 een reader samen waarin het concept Risk Perception werd uitgewerkt. Dit concept Risk Perception – vrij vertaald: risicobeleving van een individu – gaf aan marketingresearch in de V.S. een nieuwe impuls. Het bevatte een definitie van het concept Perceived Risk en een uitwerking ervan vanuit twee verschillende invalshoeken:

- de relatie tussen het concept en oriëntatiegedrag
- de relatie tussen het concept als persoonskenmerk en andere persoonskenmerken.

Het concept risicobeleving is sedertdien herhaaldelijk onderwerp van onderzoek geweest en bijna even zovele malen opnieuw gedefinieerd of omschreven:

Cox (1967)
1. Onzekerheid over het doel van de aankoop.
2. Onzekerheid over de keuzen die het best passen bij het acceptatieniveau van koopdoelen.
3. Onzekerheid over implicaties van de aankoopbeslissing.

Taylor (1974)
1. Onzekerheid over het koopresultaat: de feitelijke prestatie van het gekochte produkt in relatie tot de verwachting.
2. Onzekerheid over draagwijdte van de gevolgen van 'mismatches', zoals door Cox (2) gedefinieerd.
   Wat zijn de gevolgen en hoe ernstig zijn ze voor:
   - mijn zelfbeeld
   - mijn fysiek en psychologisch welbevinden?
   - de sociale context?

Roselius (1971) noemt als apart aspect het risico van tijdverlies voortvloeiend uit een 'miskoop'. Jacoby en Kaplan (1972) vertaalden de onzekerheden van Cox in risico's en voegden er nog aan toe: het financiële risico en het prestatierisico (prestatie voldoet niet aan de gestelde verwachtingen).

2. VERONDERSTELLINGEN OVER RISICOBELEVING EN 'STIJLEN' IN ORIËNTATIEGEDRAG ALS RESPONSIEKONSTANTES

Cunningham (1967a, 1967b) betoogde in de reader van Cox aan de hand van onderzoek dat
- risicobeleving zich concentreert binnen een segment van de totale steekproef,
- een samenhang bestaat tussen opinion leadership en risicobeleving

en konkludeert daaruit, dat risicobeleving een persoonlijkheidskenmerk is. Samen met Brody (1968) werkt hij die bevinding uit in onderzoek naar risk-perception in intermenselijke kontakten.

Deze bevindingen openen perspectieven voor nieuw onderzoek met betrekking tot masscommunicatiemedia als dragers van produktinformatie (reclame).

Dit bracht een groep uitgevers van dagbladen tot het besluit het concept risicobeleving, zoals uitgewerkt door Cunningham, te toetsen in een in Nederland te houden onderzoek.

2.1. *Model en modeleigenschappen*

De conclusies en veronderstellingen uit het onderzoek van Cunningham kunnen worden samengevat in een tweetal hypothesen:

a. *Risicobeleving als persoonskenmerk*
Het onderkennen van een risico bij kopen van een produkt van onbekend merk of het kopen in een onbekende winkel is niet alleen afhankelijk van de hoedanigheden van het te kopen produkt, doch evenzeer van de instelling van

de koper t.o.v. risico. Risicobeleving concentreert zich aldus bij een identificeerbare groep kopers.

In het vervolg van dit artikel zullen we deze veronderstelling aanduiden als de *concentratiehypothese.*

b. *Risicobeleving als gedragskenmerk*

Uit een vergelijkende analyse van personen met een relatief hoge risicobeleving met personen met een relatief lage risicobeleving leidt Cunningham af dat ook op karakteristieken van het oriëntatiegedrag de subjecten met een hogere risicobeleving op veel punten overeenkomst vertonen met het profiel van 'opinion leaders'.

In het vervolg zullen wij deze bevinding aanduiden als de *oriëntatiehypothese.*

De bevestiging van de concentratiehypothese levert een psychologische variabele op die geschikt is om oriëntatiepatronen te klassificeren onafhankelijk van de 'ruis', produkt- en omgevingskenmerken.

De implicatie hiervan voor het onderzoekmodel is, dat een longitudinaal onderzoek moet worden uitgevoerd, waarmee het besluitsvormingsproces in de tijd kan worden gevolgd. Dit heeft ook consequenties voor de op te sporen risicobeleving. Deze heeft alleen waarde als ze *stabiel* blijkt te zijn.

3. TOETSING VAN HET MODEL IN TWEE FASEN

Fase 1:

De toetsing van het door Cunningham ontwikkelde operationele criterium voor intensiteit van risicobeleving.

De verificatie van de door deze auteur aan het criterium toegekende eigenschappen.

Fase 2:

In deze fase zouden mannen en vrouwen worden ondervraagd over oriëntatiegedrag en het besluitvormingsproces met betrekking tot de aankoop van gebruiks- en verbruiksgoederen ten behoeve van de ontwikkeling van een 'taxonomie' van oriëntatiepatronen.

3.1. *Uitwerking fase 1: het onderzoek naar risicobeleving als persoonskenmerk*

Hier staan twee elementen centraal in de probleemstelling:

1. Kunnen subjecten met behulp van een eenduidig en eenvoudig operationeel criterium zodanig worden geclassificeerd, dat een consistent hogere risicobeleving met betrekking tot aankoop van uiteenlopende produkten zichtbaar kan worden gemaakt?
2. Is de aldus vastgestelde intensiteit van risicobeleving van het subject *stabiel?*

Ten behoeve van een check op stabiliteit viel de keuze op een longitudinaal onderzoek in de vorm van een test/hertest. Met een tussenpoos van ongeveer tien weken waarin twee bezoeken bij dezelfde ondervraagde worden afgelegd, zou een in grote trekken gelijkluidende vragenlijst tweemaal worden behandeld.

3.2. *Enkele technische bijzonderheden over het veldwerk en de ondervraging*

Het onderzoek werd gehouden in voorjaar en zomer van 1974. Ondervraagden waren mannen en vrouwen. Gestreefd werd naar twee steekproeven van ongeveer 600 respondenten. De eerste ondervraging vond plaats begin juni en de tweede eind augustus.

Voor het trekken van de eerste steekproef werd het 'doorloopsysteem' gehanteerd (geen herbezoeken).
Bij het tweede bezoek op hetzelfde adres eind augustus dienden wel herbezoeken te worden afgelegd.

De contactprocedures en vraaggesprekken in beide onderzoekingen leidden tot de volgende aantallen volledige gesprekken:

Tabel 1. Kontakten en response

	eerste bezoek	tweede bezoek
mannen	582	434
vrouwen	638	471

3.3. *Onderwerpen van ondervraging*

Risicobeleving bij het kopen van onbekende merken produkten of het kopen in onbekende winkels (aangeduid als 'kooprisico') werd gevraagd voor een gedetailleerd pakket van goederen en diensten:
16 verschillende niet-duurzame consumptiegoederen
20 duurzame gebruiksartikelen
2 soorten dienstverlening (brandverzekering, verzorgde vakantiereis).

Behalve naar de *intensiteit* van de risicobeleving werd ook gevraagd naar de *soorten* risico, die per produkt een rol konden spelen.
De keuze van produkten in het pakket was afgestemd op verwachtingen omtrent het soort risico dat bij de aankoop zou overheersen t.w.:
het sociale risico: reacties van vrienden en kennissen;
het prestatierisico: het falen van het produkt en het vervullen van zijn functies/prestaties;
het 'gevolgen'risico: verwachte schadelijke neveneffecten bij het gebruik van of als gevolg van het falen van het produkt.

In de afzonderlijke vragenlijsten voor mannen en vrouwen, waren niet alle produkten en diensten hetzelfde, omdat bepaalde produkten specifiek tot het

'beslissingsdomein' van de man en andere tot dat van de vrouw moeten worden gerekend. In dat geval werden substituten ingevoerd waarvan een zekere analogie in intensiteit een soort kooprisico kon worden aangenomen.

In de vraagstelling werd rekening gehouden met de marketing- en concurrentiekarakteristiek van het produkt. Naar het risico van een onbekend merk werd gevraagd bij die produkten, waarvan het merk inzet is van marketing en concurrentie. Naar het risico van het kopen in onbekende winkels werd gevraagd bij die artikelen, waar de concurrentie zich concentreert rond activiteiten op het verkooppunt (bijvoorbeeld meubelen, gordijnen etc.).

Het bestond uit referentiekader voor deze toets, die de betrokkenheid van het subject bij het kopen van genoemde artikelen moesten bepalen enerzijds, en vragen zoals gebruikt bij het opsporen van opinieleiderschap anderzijds. (belangstelling voor reklame over...)

4. UITKOMSTEN VAN HET ONDERZOEK EN ANALYSE

Indien de risicobeleving bij de ondervraagde groep gestructureerd zou zijn conform de hypothese van Cunningham, dan zou deze zichtbaar kunnen worden gemaakt door een scalogram, waarin *alle* produkten figureren als een homogene reeks items met een (monotoon) stijgende moeilijkheidsgraad.

Rekening houdend met het (ook hier) optreden van 'ruis', werd het juister geoordeeld om het schaalcriterium holomorfie te kiezen in plaats van het zeer strenge Guttmancriterium voor homogeniteit.

Het materiaal werd op grond van deze overwegingen onderworpen aan de Mokken-schaalanalyse.

Deze analyse leverde niet het beoogde resultaat. De items (produkten, diensten) werden gegroepeerd tot een lange, zwakke schaal (28 items: schaalcoëfficiënt: .40) en een korte, zwakke schaal (8 items schaalcoëfficiënt: .35).

In tabel 2 is een overzicht opgenomen van uitkomsten van de ondervraging in de steekproef mannen: het bevat een opsomming van produkten en diensten waarover het 'kooprisico' werd gevraagd. De linkerkolom bevat het percentage van de somscore 'tamelijk riskant en zeer riskant'.

Uit het feit dat de scalogramanalyses twee vrij zwakke schalen genereerden in plaats van één sterke, kon worden geconcludeerd dat de concentratiehypothese van Cunningham, vertaald in deze eis, vatbaar was voor relativering.

Het aanleggen van een hogere drempelwaarde voor toelating van items op de schaal (schaalcoëfficiënt Loevinger $H \geqslant .5$) had tot gevolg de selectie van 16 items uit een totaal van 38.

Uit de opsomming van items kon worden opgemaakt dat bij de selectie van items met een laag waargenomen risico de levensmiddelen, dranken en cosme-

Tabel 2. Ontwikkeling van een 'algemene' kooprisicoschaal. Vragenlijst 'mannen', steekproef: (n = 582)

Item nummer	Omschrijving	Kopen onbekend merk/winkel tamelijk + zeer riskant	item coëfficiënt in Mokken-schaal selectie bij Loev. H ≥ .50
34	papieren zakdoekjes	5%	.3677
10	staande schemerlamp	8%	.5626
13	stropdassen	8%	.4986
32	bad-/doucheschuim	15%	
11	ondergoed	16%	.4273
35	pre-shave	16%	
14	bloembollen	17%	
37	haarlotion	17%	
12	overhemden	20%	.4998
26	updranken	21%	
27	sinaasappelsap	22%	
36	after-shave	23%	
33	deodorant	24%	
20	bovenkleding	26%	.4378
7	overgordijnen	29%	.4504
23	margarine	31%	
24	sherry	32%	
29	wasmiddel voor de grote was	37%	
19	bankstel	37%	
25	bier	38%	
28	soep in blik	40%	
18	broodrooster	43%	.4753
16	fiets	45%	
6	vast tapijt	48%	
30	insectenbestrijdingsmiddel	50%	
38	vers vlees	50%	
2	elektrisch koffiezetapparaat	59%	.5144
15	elektrisch scheerapparaat	65%	.5279
17	elektrische boormachine	66%	
5	losse platenspeler	67%	.5019
4	zwartwit-TV	69%	.5441
3	elektrische naaimachine	69%	.5127
9	band-/cassetterecorder	75%	.5351
1	personenauto	81%	.4961
22	vakantiereis	82%	.6586
21	brandverzekering	82%	
8	kleuren-TV	85%	

tica systematisch waren vermeden (zie tabel 2, rechterkolom).

Een herhaling van deze analyse in de steekproef vrouwen leverde een geheel analoog resultaat.

Dit verschijnsel kon op drie verschillende manieren worden verklaard:

- een zeer geringe betrokkenheid van de ondervraagden bij het 'kooprisico' van deze groep artikelen;
- de interventie van een nieuwe variabele in het scorepatroon, aan te duiden als een zekere mate van gespecialiseerde kennis omtrent bepaalde groepen produkten (bijvoorbeeld woninginrichting en kleding);

– de gevoeligheid van de Mokken-schaalanalyse voor het optreden van series items met een stijging van de populariteitswaarde in *kleine* stappen (kleine waarden $Y_i - Y_j$).

  Aangezien met name bij levensmiddelen, dranken en cosmetica deze kleine waarden van $Y_i - Y_j$ veelvuldig optraden, vertoont de selectie van items uit die groep via deze methode veel overeenkomst met 'toevalstrekking' (Jansen, 1982).

Voor het kiezen uit mogelijke verklaringen was een vervolganalyse noodzakelijk. Besloten werd om de samenhang tussen risicobeleving en specialisatie nader te onderzoeken.

4.1. *Analyse van het materiaal op specialisatie van risicobeleving*

De ondervraagde mannen en vrouwen hadden zich niet uitgesproken over specialistische kennis, hobbies etc., zodat het optreden hiervan en de invloed op de structuur van risicobeleving niet feitelijk kon worden aangetoond.

Een multivariate analyse op scores risicobeleving van de afzonderlijke produkten/diensten zou indicaties over de samenhang tussen specialisatie en risicobeleving kunnen opleveren.

Het materiaal uit de steekproef mannen en dat uit de steekproef vrouwen werd afzonderlijk aan een factoranalyse onderworpen. Hierin fungeerden *ondervraagden* als *'subjecten'* en de produkten/diensten als *operatoren.*

De partiële correlaties van de risicoscores vormden de basis van een analyse van principale componenten. De rotatie van het assenstelsel geschiedde aan de hand van VARIMAX-criterium.

Op grond van het verloop van gevonden eigenwaarden werd beloten de oplossing met drie factoren nader uit te werken. De groepering van operatoren (produkten/diensten) met ladingen $\geqslant .50$ is opgenomen in het overzicht van tabel 3.

Zowel analyse van de steekproef mannen als die van vrouwen leverde bij een driefactoroplossing een totale 'verklaarde' variantie op van $\pm 50\%$. In beide gevallen nam de eerste factor meer dan de helft hiervan voor zijn rekening.

In beide gevallen omvat de eerste factor niet-duurzame consumptiegoederen met een overwegend laag 'kooprisico'. Nadere analyse leerde dat prestatierisico hier als soort risico overheerste. In enkele gevallen (bier en sherry) had het sociaal risico de overhand.

Op de derde factor treft men textiel en woninginrichting aan met een laag tot middelhoog kooprisico. Ook hier overweegt prestatierisico.

De tweede factor wordt verzamelpunt van duurzame gebruiksgoederen met

Tabel 3. Verdeling van produkten over factoren (clusters) naar analogie van intensiteit van risicobeleving (kopen onbekend merk)
Produkten in volgorde van hoogte van de lading op de desbetreffende factoren (alleen lading ≥.50)

	Mannen	Vrouwen
Factor I	after-shave, deodorant, pre-shave, sinaasappelsap, bad-doucheschuim, updranken, haarlotion, bier, sherry, soep, margarine, wasmiddel, tissues.	sinaasappelsap, bier, updranken, sherry, soep, deodorant, lipstick, bad-/doucheschuim, maandverband, wasmiddel, margarine, kleurshampoo, tissues.
Factor II	zwartwit-TV, band-/cassetterecorder, kleuren-TV, elektr.naaimachine, personenauto, elektr.koffiezetapparaat, idem scheerapparaat, boormachine, broodrooster.	zwartwit-TV, band-/cassetterecorder, kleuren-TV, elektr.naaimachine, idem platenspeler, koffiezetapparaat, personenauto, elektr. boormachine, broodrooster, scheerapparaat.
Factor III	overhemden, stropdassen, ondergoed, bovenkleding, overgordijnen, schemerlamp, bankstel, vast tapijt.	schemerlamp, jurken, truitjes, gordijnen, bloembollen, corsetterie, bankstel.

een hoog tot zeer hoog kooprisico. In de meeste gevallen werd aan prestatierisico en gevolgenrisico een nagenoeg even grote plaats toegekend.

Het was erg verleidelijk om de gevonden latente structuur te bestempelen als specialisatie, doch de opsomming van neveneigenschappen van de afzonderlijke groepen (het optreden van error scores, kleine waarden van $Y_i - Y_j$) ontkrachtte de alternatieve verklaringen niet voldoende.

4.2. *Specialisatie en concentratie*

Als laatste stap in de analyse van de kwantitatieve data werd besloten de afzonderlijke groepen van produkten en diensten die de factoranalyse had voortgebracht, aan een schaalanalyse te onderwerpen, op zoek naar steun voor de veronderstelling.
- risicobeleving hangt samen met een specifieke kennis en ervaring van de consument;
- binnen de specifieke kennisgebieden bepaalt de persoonsstructuur de intensiteit van de risicobeleving.

De factor II (duurzame gebruiksartikelen) leverde zowel in de mannen- als in de vrouwensteekproef een 'sterke' schaal op met schaalcoëfficiënten .5062 (mannen) en.5613 (vrouwen).

De factoren I en III brachten het in geen van beide steekproeven verder dan een 'medium'-schaal.

Kortheidshalve geven we daarom alleen de resultaten op factor II (duurzame

gebruiksartikelen) weer in tabel 4.

Tabel 4. 'Mokken'-schalen risicobeleving duurzame gebruiksartikelen

Item nummer	Omschrijving	Vragenlijst vrouwen item coëff.	Vragenlijst mannen item coëff.
18	broodrooster	.6053	.5564
15	elektrisch scheerapp. voor vrouwen	.5322	
15	elektr. scheerapp.		.5027
2	elektr.koffiezet-apparaat	.5272	.5096
5	losse platenspeler	.5449	.4937
	elektr. boormachine	.5082	.3816
3	elektr.naaimachine	.5479	.4967
4	zwartwit-TV	.5881	.5243
9	band-/cassetterecorder	.5932	.5425
1	personenauto	.5079	.4851
8	kleuren-TV	.7277	.6434
Schaalcoëfficiënt (Loev. H)		.5613	.5062

5. DE STABILITEIT VAN RISICOBELEVING

Aan de orde was thans de eis van stabiliteit van de gevonden scores, zoals geformuleerd in paragraaf 2.1. De in het tweede onderzoek door dezelfde respondenten gegeven antwoorden werden met die uit het eerste vergeleken. Van centraal belang was daarbij het optreden van verschillen in de scores risicobeleving tussen tijdstip 1 en tijdstip 2.

Deze verschillen bleken aanzienlijk te zijn. Met name bij de produkten met hoge score risicobeleving behaalde slechts 40%-50% van de ondervraagden in beide onderzoekingen dezelfde scores.

Zeer opvallend was het resultaat dat bij het herbezoek op 29 van de 31 items die uiteindelijk in een Mokken-schaal een plaats hadden gevonden, de tweede score hoger uitviel dan de eerste. Uitzonderingen vormden het broodrooster en de wasmachine, waarvan de eerste scores hoger uitvielen.

Verschillen in de omvang van de verschuiving hingen alleen samen met de hoogte van het initiële 'waargenomen risico' en waren derhalve over de hele linie gelijkwaardig.

Kennelijk is er sprake van een consistente 'bias' in de test/hertest-procedure, die losstaat van andere in het onderzoek geregistreerde. Uit een voortgezette secundaire analyse bleek, dat alleen de koopintentie t.a.v. het beoordeelde produkt enige – niet significante invloed had op de mate waarin bij de hertest het risico hoger werd geschat. Indien tijdens het tweede vraaggesprek het geplande tijdstip van aankoop merkbaar naderbij was gekomen, viel de risicoschatting systematisch hoger uit.

6. VERKENNING VAN RISICOBELEVING EN BESCHRIJVING VAN KOOPBESLISSINGEN

De onverwachte en niet eenduidige verklaring in het gedrag van de 'operatoren' kooprisico en risicocategorieën maakten de uitvoering van de tweede fase van het onderzoek zinloos.

Wel was het aanleiding tot het uitvoeren van een 'kwalitatief onderzoek' ter nadere verkenning van de invulling, die respondenten aan het begrip risico zouden geven.

6.1. *Eerste aanvullende verkenning; groepsdiscussies*

In eerste instantie besloot men tot het houden van vier groepsdiscussies. Deelnemers aan de discussies werden 'gescreend' aan de hand van een vragenlijst.

Achtereenvolgens werden geselecteerd:
- een groep vrouwen met hoge risicobeleving;
- een groep vrouwen met lage risicobeleving;
- een gemengde groep vrouwen (sommige met hoge, andere met lage risicobeleving);
- een gemengde groep mannen (sommige met hoge, andere met lage risicobeleving).

De deelnemers aan de groepsdiscussies hanteerden het begrip risico op een wijze die sterk deed denken aan de variëteiten van perceived risk volgens Jacoby en Kaplan (1972).

De inhoudsomschrijving van het begrip risico is in het algemeen ontleend aan besluitvormingsprocessen ten aanzien van de aanschaf van dure gebruiksartikelen (gezinsinvesteringen). Onvermijdelijk werd daarmee het begrip risico geplaatst in het kader van de keuze van de leverancier, *meer* dan in het kader van de keuze van het merk.

Het besprokene tijdens de groepsdiscussie vestigde de aandacht op strategieën van risicobeperking door middel van delegatie van risico aan derden. Centraal stond hierbij de rolverdeling in het beslissingsproces tussen de echtelieden (inbreng van expertise, het dragen van verantwoordelijkheid, inspraak in beslissing). Tevens werd de aandacht gevestigd op de mogelijkheden tot delegatie van verantwoordelijkheid aan derden (vrienden en kennissen, de verkoper).

Voor wat betreft de rolverdeling tussen echtelieden leidden de groepsdiscussies tot een aanzet voor een typologie die zich tussen twee extremen beweegt:
- het gehele beslissingsproces wordt door man en vrouw 'samen' doorlopen;
- elk van de partners is 'expert' op eigen terrein. Inmenging van de ander in het eigen beslissingsdomein wordt niet geduld.

Tussen deze beide extreme rolpatronen treft men diverse mengvormen aan.

Aangezien het tegenwoordig bij bijna alle grote huishoudelijke apparatuur gaat om vervangingsaankopen, speelt bij koopbeslissingen het ongemak, ont-

staan door het defect raken, onherstelbaar zijn van het oude apparaat in de koopbeslissing als 'tijdrisico' mee.
Dit ongemak onthult een zekere *onmisbaarheid* van deze apparatuur, die het zorgvuldig afwegen van alternatieven en kiezen in de koopbeslissing dreigt te doorkruisen.

Voorbeelden:
- de kleuren-TV geeft de geest een paar dagen vóór een belangrijke voetbalwedstrijd; het defect raken van de wasmachine, etc.

Newall (1977) kent in zijn studie over inkoopgedrag in industriële ondernemingen aan deze *onmisbaarheid* van het te kopen artikel een aparte plaats toe: 'the degree of product essentiality'.

Ook Newall komt tot de conclusie dat de onmisbaarheid van het artikel aanwijsbaar van invloed is op de tijdsduur die het nemen van een koopbeslissing in beslag neemt.

6.2. *Tweede aanvullende verkenning; vrije interviews met echtparen*

De uit de groepsdiscussies naar voren gekomen tendens naar het delegeren van verantwoordelijkheid en risico in het kader van rolrelaties tussen partners was aanleiding om het verkenningsprogramma uit te breiden met en zestal vrije interviews, waarin simultaan de man èn de vrouw in het gezin werden ondervraagd.

Het selectiecriterium voor de te ondervragen echtparen was de recente aanschaf van een wasmachine. In de vraaggesprekken viel de nadruk op het verloop van het besluitvormingsproces en de visie van de echtelieden op hun eigen aandeel en hun rol in dit proces. De bevindingen uit deze interviews leverden in grote trekken een bevestiging van de resultaten van de groepsdiscussies op.

Een belangrijk nieuw aspect was dat het risico dat met de koopbeslissing is gemoeid, in de loop van het proces voortdurend van inhoud wisselt. Deze wisseling hangt samen met een verandering van de 'focus'; het 'zoeklicht' op onderdelen van het keuzedilemma in de loop van het besluitvormingsproces zelf. (Zie ook de 'evoked sets' van Howard & Sheth (1969))

Uit de verslagen van de vrije interviews tekende zich een stereotiep verloop van koopbeslissingen af:
- in het beginstadium van het beslissingsproces werd de meeste aandacht besteed aan de keuze van het *merk* resp. *type*. Het oriëntatiegedrag en de verdeling van taken daarin zijn primair daarop gericht.

Nadat een of meer alternatieven uit dit keuzeproces naar voren zijn gekomen treedt een verandering in de 'focus' van de besluitvorming op: aan de orde is dan de keuze van de *winkel*.

Het 'winkelrisico' is opgebouwd uit de elementen:
- financieel risico: verschillende winkels vragen voor dezelfde merkentypen wasmachines verschillende prijzen;
- prestatierisico. Een aspect van goede service is controle op het te leveren artikel en een deugdelijke installatie. Een ander aspect is coulance indien de nieuwe machine niet feilloos blijkt te functioneren: de machine wordt vlot

vervangen; het defect wordt vlot verholpen.
- Een gevolgen- resp. 'tijd'-risico. Zo *goedkoop* en *snel* mogelijk verhelpen van mankementen helpt deze risico's verkleinen.

Uit het aantal keren dat de ondervraagden tijdens de bespreking van de koopbeslissing ten aanzien van de wasmachine refereerden aan andere koopbeslissingen (auto, TV-toestel, koelkast, etc.) bleek, dat dit stereotiepe verloop van het beslissingsproces en alle daarop voorkomende varianten niet exclusief voor de aankoop van de wasmachine geldt, doch zich vermoedelijk uitstrekt over vele andere aankopen van duurzame gebruiksartikelen die het karakter dragen van een 'gezinsinvestering'.

7. CONCLUSIE EN STELLINGEN

7.1. Toepassing van de theorie en de operationalisaties van risicobeleving als persoonskenmerk, zoals ontwikkeld door Cunningham, levert complicaties op.

De validiteit van de operationalisatie van het concept kooprisico of risicobeleving, zoals uitgewerkt in de vraag of en *in welke mate* het subject het kopen van produkten van een onbekend merk of in een onbekende winkel riskant vindt, is van twijfelachtig gehalte. Aspirant-kopers zijn meer gepreoccupeerd met andere aspecten van het risico, dan met de omvang ervan.

7.2. De besluitvorming ten aanzien van het kopen van produkten van een onbekend merk verloopt – al naar de aard van het produkt en de daarmee gemoeide uitgaven – op zeer uiteenlopende wijze.

Initiële of vervangingsaankopen, die het karakter dragen van een 'gezinsinvestering', gaan gepaard met een besluitvormingsproces dat in stadia verloopt.

Veelal kunnen in dat proces ten minste twee stadia worden onderkend:
- de totstandkoming van een 'evoked set' rond de keuze van merk en type;
- de totstandkoming van een tweede 'evoked set' rond de keuze van de leverancier (winkeltype, reputatie van dealer).

De ontwikkeling van elk van deze 'evoked sets' is gekoppeld aan een specifieke risicobeleving; de aard van het waargenomen risico is specifiek verbonden aan dat onderdeel van het beslissingsproces resp. de mogelijkheden van beperking ervan.

7.3. De individuele taxatie van het risico als (relatief klein of groot) is mede afhankelijk van de mate waarin het individu geheel zelfstandig of in gedeelde verantwoordelijkheid met andern over de aankoop beslist.

Het verloop van een besluitvormingsproces ten aanzien van gezinsinvesteringen is sterk afhankelijk van het in het gezin geldende rolpatroon tussen de echtelieden. Dit rollenpatroon bepaalt of en in welke mate verantwoordelijkheid wordt gedelegeerd dan wel gedeeld.

Dat deze rolverdeling langs deze weg van invloed zal zijn op de risicobeleving van het individu (omvang van het waargenomen risico) staat buiten twijfel.

7.4. Naast de delegatie van verantwoordelijkheid en de daaraan gekoppelde 'spreiding' van het risico tussen echtelieden treedt ook op de poging tot delegatie van verantwoordelijkheid aan derden met hetzelfde oogmerk: het verminderen van het risico.

In veel besluitvormingsprocesen is het raadplegen van de verkoper en het vragen van advies aan vrienden en kennissen te beschouwen als zo'n poging. Via de raadpleging wordt getracht de betrokkenen te 'compromitteren'.

7.5. Een belangrijke indirecte factor die de mate van het waargenomen risico zal beïnvloeden is de situatie waarin een als onmisbaar onderkend gebruiksartikel moet worden vervangen. Het onverwachts en onherstelbaar defect raken van een als onmisbaar onderkend gebruiksartikel brengt het gehele besluitvormingsproces ten aanzien van de vervanging in een hogere versnelling.

7.6. De door Cunningham (1967) gehanteerde indeling van risico's: (sociaal risico, prestatie- en gevolgenrisico) blijkt te simplistisch voor de uitwerking in een onderzoek. Volgens discussies en vrije interviews refereren de gesprekspartners zelden in directe zin aan *risico;* zij brengen het risico indirect ter sprake in de beschrijving van strategieën tot vermindering ervan.

Bij een poging tot terugvertaling van deze strategieën in soorten risico komt men tot veel meer soorten risico dan Cunningham onderscheidt.

Jacoby en Kaplan blijken met hun indeling van risico's

- financieel risico
- prestatierisico
- fysiek risico
- sociaal risico.

de werkelijkheid dichter te benaderen.

Ook de door Roselius (1971) beschreven factor *tijdsrisico* manifesteert zich duidelijk in het onderzoek.

7.7. Het psychologisch risico dat verschillende auteurs nader definiëren als risico van schade aan het zelfbeeld dat kan ontstaan door een verkeerde merken- en produktkeuze bleek geen geanimeerd gesprekstheme in groepsdiscussies en vrije interviews.

In een kwantitatief onderzoek dook een variant van dit risico op bij de keuze van het merk bier of sherry. Voor deze produkten erkenden de ondervraagden het bestaan van een sociaal risico, omschreven als: 'wat vrienden en kennissen ervan denken'.

7.8. De onderkenning van risico's bij de aankoop van onbekende merken en in onbekende winkels lijkt samen te hangen met de beschikbaarheid van specialistische kennis. (zie § 4.1)

De ondervraging van mannen en vrouwen over risico's die optreden bij een groot aantal uiteenlopende produkten en diensten geeft een tendens aan tot specialisatie. Deze specialisatie echter kan méér dan één oorzaak hebben. *Niet* is vastgesteld dat bij deze indeling de aanwezigheid van specialistische kennis of bijzondere interesse de voornaamste oorzaak is.

7.9. De concentratiehypothese lijkt wel houdbaar, maar het *fenomeen* wordt het duidelijkst zichtbaar bij koopbeslissingen t.a.v. duurzame gebruiksartikelen.

7.10. Het afnemen van een vraaggesprek waarin met de monotonie van een uniforme vraagstelling respondenten de mate van risico, gemoeid met het kopen van onbekende merken/in onbekende winkels over grote aantallen produkten en diensten moeten taxeren, leidt tot beïnvloeding van de houding van de ondervraagden ten aanzien van risicobeleving.

Er zijn aanwijzingen dat zo'n vraaggesprek het probleem van het optreden van risico's actualiseert.

Bij een herhaling van het vraaggesprek tien weken na dato scoort de respondent consistent hoger op de schalen van intensiteit van de risicobeleving.

Dit impliceert dat de test/hertestmethode als hulpmiddel om stabiliteit van risicobeleving als attitude te bepalen, ongeschikt is.

7.11. Het naderen van het geplande tijdstip van aankoop was de enige referentievariabele die enige samenhang vertoonde met een intensivering van de risicobeleving. Deze samenhang was echter niet significant.

7.12. De inventarisatie van soorten risico in het kwalitatieve onderzoek alsmede de gebleken invloed van omstandigheden waaronder een koopbeslissing wordt genomen en stijl van beslissen wordt toegepast, maakt duidelijk, dat risicobeleving een sterk aan het produkt en aan de omstandigheden van de koopbeslissing gebonden verschijnsel is.

Al naar gelang de aard van het produkt en de markt waartoe het behoort, alsmede de omstandigheden waaronder wordt beslist, manifesteert kooprisico zich in vormen die kunnen variëren van het (gemakkelijk) formuleren van een eenduidig risico tot de veelkoppige draak die zich gedurende een lang proces van besluitvorming op verschillende momenten in verschillende gedaanten aandient.

7.13. De in het onderzoekprogramma vervatte volgorde van kwalitatief verkennend onderzoek, dat volgt op een uitvoerig kwantitatief toetsenonderzoek, is ongebruikelijk. In het onderhavige geval was deze volgorde een uitvloeisel van het feit dat een kwalitatief en kwantitatief onderzoekprogramma uit de Verenigde Staten voorhanden was en voldoende werd geoordeeld als basis voor operationalisering van een 'gesloten' vragenlijst.

De gang van zaken rond dit onderzoek vormt voor gevallen van deze een pleidooi om – in weerwil van de beschikbaarheid van onderzoekresultaten uit het buitenland – een verkennend programma van onderzoek vooraf te laten gaan aan een omvangrijk beschrijvend of toetsend onderzoek.

BIBLIOGRAFIE

Brody, Robert P. and Cunningham, Scott M., 1968. *Personality Variables and the Consumer Decision Process, Journal of Marketing Research, 5* pp. 50-56.

Cox, Donald F. (ed.) 1967 *Risk Taking and Information Handling in Consumer Behaviour, Introduction* Harvard University Press, pp. 1-19.

Cunningham, Scott M. (1967a), (1967) *The Major Dimension of Perceived Risk* in: D. Cox (ed.), *Risk Taking and Information Handling in Consumer Behavior,* Harvard University Press, pp. 82-108.
Cunningham, Scott M. (1967b), 1967. *'Perceived Risk as a Factor in Informal Consumer Communications,* in: D. Cox (ed.), *Risk Taking and Information Handling in Consumer Behavior.* Harvard Universtity Press, pp. 265-288.
Howard, J.A. en Sheth, Jagdish N. 1969. *The Theory of Buyer Behavior* Wiley, New York.
Jacoby, Jacob en Kaplan, Leon B., 1972. *'The Components of Perceived Risk'.* Proceedings of the 3rd American conference of the Association for Consumer Research, pp. 382-393.
Jansen, Paul G.W., 1982. *De onbruikbaarheid van de Mokken-schaalanalyse,* in: *Tijdschrift voor Onderwijsresearch,* 7:1, pp. 11-24.
Katz, Elihu and Lazarsfeld, Paul F., 1955 *Personel Influence,* Glencoe, Illinois: The Free Press.
Molenaar, Ivo W., 1982. *De beperkte bruikbaarheid van Jansen's kritiek,* in: *Tijdschrift voor Onderwijs Research,* 7:1, pp. 25-30.
Dr. John Newall, 1977. *Industrial Buyer Behaviour, European Journal of Marketing,* vol. 11. 3-1977, pp. 165-211.
Roselius, Ted, 1971. *Consumer ranking of risk reduction methods,* Journal of Marketing, 35, pp. 56-61.
Taylor, James W., 1974, *The Role of Risk in Consumer Behavior Journal of Marketing,* 38, pp. 54-60.

13. Panels in de econometrie

T.J. WANSBEEK*

SAMENVATTING

De analyse van panelgegevens (tijdreeksen van dwarsdoorsnedegegevens) krijgt in de econometrie in toenemende mate aandacht. In dit artikel wordt daarvan een inleidend overzicht gegeven, met het zwaartepunt op het regressiemodel met aparte tijd- en waarnemingseenheidseffecten, en op de problemen die door selectieve uitval ontstaan.

ABSTRACT

The analysis of panel-data (time-series of cross-sections) is gaining increasing attention in econometrics. This paper gives an introductory overview of the subject, emphasizing the multiple regression model with time and observation unit effects, and the problems connected with selective attrition.

1. INLEIDING

In toenemende mate bestaat er in de econometrie belangstelling voor het analyseren van panelgegevens. Anders dan in de marktkunde, waar het gebruik van dit soort data een lange geschiedenis heeft, is de belangstelling in de econometrie van recentere datum, maar in vrij korte tijd een bloeiende tak van zowel theoretisch als empirisch onderzoek geworden. Omdat deze ontwikkeling grotendeels onafhankelijk van de marktkunde heeft plaatsgevonden, lijkt het de moeite waard om in dit artikel in te gaan op de manier waarop de econometrie met panels omspringt, zodat een marktonderzoeker weet wat hij kan verwachten van de econometrische literatuur op dit punt.

Wie zich hiervan – buiten dit artikel en de daarin vervatte referenties – een oordeel wil vormen kan terecht bij een aantal econometrische leerboeken, met name Kmenta (1971), Maddala (1977), en vooral het voortreffelijke boek van Judge e.a. (1980), en bij de 'proceedings' van een in Parijs gehouden conferentie over de econometrie van panel-data (Mazodier (1978) – zie Kapteyn en Wansbeek (1978) voor een kort verslag van deze conferentie en een inleidend overzicht van de panel-econometrie).

Het is nuttig om allereerst vast te stellen wat we onder 'panelgegevens' verstaan. We volgen in dit artikel het spraakgebruik van de econometrie, waar

* Hoofdafdeling Statistische Methoden, Centraal Bureau voor de Statistiek. De in dit rapport weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek. De auteur dankt prof. dr. ir. A. Kapteyn, prof. dr. W.J. Keller, prof. dr. M.T.G. Meulenberg, drs. F.J.R. van de Pol en drs. H. van de Stadt voor hun commentaar op een eerdere versie van dit artikel.

men een ruim begrip hanteert en van panelgegevens spreekt als het gaat om een gegevensbestand met informatie over een aantal waarnemingseenheden (zoals individuen, bedrijven, provincies) op een aantal tijdstippen. Deze omschrijving dekt een veelheid aan manieren van verzamelen. Het kan gaan om gegevens die afkomstig zijn van een panel zoals bij het marktonderzoek wordt gebruikt, dat wil zeggen een groep individuen of gezinnen die regelmatig of zelfs doorlopend worden bemonsterd. Ook kan het gaan om door de onderzoeker bijeengebrachte, gewoonlijk voor een ander doel verzamelde tijdreeksen voor een aantal waarnemingseenheden, zoals provincies of bedrijven, of om de uitkomsten van een ten minste eenmaal bij dezelfde steekproef herhaald opinie-onderzoek. Wat al deze bestanden gemeen hebben is, dat ze bestaan uit gegevens die via twee ingangen worden gekarakteriseerd, namelijk tijdstip en waarnemingseenheid.

In het meest ideale geval is het panelbestand *volledig,* dat wil zeggen dat voor alle waarnemingseenheden alle variabelen op alle tijdstippen zijn gemeten. Volledige panelbestanden zijn eenvoudiger te analyseren dan *onvolledige* panelbestanden, waarbij *niet* alle waarnemingseenheden op alle tijdstippen zijn waargenomen. In de econometrie is tot nu toe de meeste aandacht besteed aan volledige panelbestanden, omdat de belangstelling oorspronkelijk vooral uitging naar de analyse van gecombineerde tijdreeksen van een groep bedrijven of provincies; de onderzoeker zorgt, bij het verzamelen van zijn gegevens, er wel voor een volledig bestand te hebben.

Gedurende de laatste tien jaar is de econometrie zich meer en meer gaan bezighouden met gegevensbestanden die afkomstig zijn van grote personen- en gezinsenquêtes, en daarmee is er ook belangstelling ontstaan voor onvolledige panelbestanden: wanneer een grote groep individuen een aantal malen wordt benaderd voor een enquête, dan zullen, naast de initiële non-respons, naarmate de tijd verloopt, steeds meer geënquêteerden het laten afweten ('attrition', afslijting); het panelbestand gaat steeds meer lege plekken vertonen. Onvolledigheid van een panelbestand kan ook met opzet teweeg worden gebracht, als het panelonderzoek op roterende basis is opgezet, en een deel van het panel (bijvoorbeeld de helft of een kwart) in iedere golf wordt vervangen door nieuwe respondenten. Ook al is er geen non-respons, dan nog is, als men na een aantal golven van het onderzoek een gegevensbestand maakt, dat bestand onvolledig omdat niet alle waarnemingseenheden op alle tijdstippen zijn bemonsterd.

De econometrie van panelgegevens is vooral tot bloei gekomen in de Verenigde Staten. Daar worden, sinds een jaar of vijftien, regelmatig grote panelonderzoeken op touw gezet om experimenteel de invloed op het gedrag van gezinnen en individuen na te gaan van allerlei mogelijke sociaal-economische maatregelen. Het gaat hierbij om micro-economisch onderzoek van dynamische processen die alleen op een individueel niveau kunnen worden geanalyseerd.

Ook in Nederland bestaat een groeiende belangstelling voor panels bij sociaal-economisch onderzoek. Zo loopt sinds maart 1980 bij het CBS het zogenaamde Inkomenswaarderingspanel (IWP), waarbij ongeveer 1000 gezinnen ieder jaar worden ondervraagd over hun inkomen en de waardering ervan. Een theoretisch verantwoorde analyse (zie bijvoorbeeld Kapteyn, 1977,

Kapteyn en Wansbeek, 1982) is pas mogelijk als van gezinnen informatie over tenminste twee jaar beschikbaar is, en deze wordt door het IWP geleverd. Het is overigens de bedoeling dat het IWP in de nabije toekomst zal worden ingebouwd in een meer omvattend en breder opgezet Sociaal-Economisch Panel, dat nog meer beleidsrelevant onderzoek mogelijk moet maken. Hierbij wordt gedacht aan onderzoek naar bijvoorbeeld inkomensmobiliteit (zijn het steeds dezelfde mensen die het minimumloon verdienen, of wisselt deze groep sterk?), verdeling premiedruk (zijn het altijd dezelfde mensen die sociale-verzekeringspremies betalen en anderen die uitkeringen ontvangen?) en ongelijkheid in levensduurinkomen. Voor al deze onderzoeksgebieden geldt dat de economische theorie dynamische verbanden postuleert tussen op individueel niveau waargenomen variabelen. Wanneer men deze verbanden tracht te kwantificeren op grond van een enkele doorsnede of een tijdreeks van aggregaten, dan mist men niet alleen de rijke variatie die een panel biedt, maar zullen in het algemeen de gehanteerde schattingen worden afgeleid uit een onjuist gespecificeerd model en dientengevolge tot onjuiste conclusies leiden.

Om een indruk te geven van de econometrie van de panelgegevens zal in dit artikel nogal uitgebreid worden ingegaan op een enkel, eenvoudig model, namelijk het model met een samengestelde constante term. Over dit model is niet alleen het meeste geschreven, het laat zich ook goed gebruiken om allerlei typisch econometrische problemen te illustreren. Aan dit model zijn de volgende twee paragrafen gewijd. Een aantal andere modellen, voor een deel uitbreidingen van dit eenvoudige model, worden in paragraaf 4 besproken. De behandeling tot dan toe richt zich op volledige panelbestanden met hun eenvoudige structuren. In paragraaf 5 worden een aantal andere modellen kort besproken. Onvolledige panelbestanden komen in paragraaf 6 aan de orde. In paragraaf 7 worden een paar afsluitende opmerkingen gemaakt.

2. HET MODEL MET SAMENGESTELDE CONSTANTE TERM

Bezie het gewone lineaire-regressiemodel, waarin een afhankelijke variabele (y) wordt verklaard door onafhankelijke variabelen (samen de kolomvector x vormend). De index n ($n=1, \ldots, N$) duidt de waarnemingseenheid aan, en t ($t=1, \ldots, T$) de tijd. Laat ε de storingsterm zijn en β een kolomvector van coëfficiënten, dan luidt het gewone regressiemodel, voor waarnemingseenheid n op tijdstip t:

$$y_{nt} = x'_{nt}\beta + \varepsilon_{nt}. \tag{2.1}$$

De meest gebruikelijke veronderstelling over de storingsterm is dat ε_{nt} een stochastische grootheid is met verwachting nul en een constante variantie σ_ε^2, dat wil zeggen hetzelfde voor alle n en alle t. Een van de elementen van x_{nt} is identiek gelijk aan 1, en de waarde van het bijbehorende element van β (stel het eerste, β_1) geeft dus de grootte van de constante term aan.

Het voornaamste doel dat een econometrist zich stelt is om de elementen van de vector β op een of andere 'goede' manier te schatten. Een eerste voordeel van panel-data blijkt meteen: vergelijking (2.1) kan worden geschat op grond van NT waarnemingen, wat betrouwbaarder resultaten oplevert dan een tijdreeks (T waarnemingen) of een doorsnede (N waarnemingen). Zo on-

derzoeken, in een klassiek geworden artikel op het gebied van het 'poolen' van tijdreeks- en doorsnedegegevens, Balestra en Nerlove (1966) de vraag naar aardgas in de Verenigde Staten niet aan de hand van geaggregeerde cijfers voor het hele land, maar via aparte tijdreeksen (1950 – 1962, dus $T = 13$) voor de afzonderlijke staten ($N = 36$; niet in alle staten is ieder jaar aardgas gebruikt). Zodoende zijn dus $13 \times 36 = 468$ waarnemingen beschikbaar.

Specificatie (2.1) is in een belangrijk opzicht restrictief. De coëfficiëntenvector β heeft geen index n of t en is dus stilzwijgend constant verondersteld. Dat kan een aanvaardbare hypothese zijn – denk aan het voorbeeld van de vraag naar aardgas: het is wellicht redelijk om te veronderstellen dat de prijsgevoeligheid van die vraag niet varieert over de tijd of over de verschillende staten. Met de constante term ligt het anders. Het aardgasverbruik zal per staat systematisch op een verschillend niveau liggen, bijvoorbeeld als gevolg van een verschil in geografische ligging. Om met het gezamenlijke effect van al dit soort over de tijd constante, niet expliciet opgenomen variabelen rekening te houden in (2.1) ligt een aangepast model voor de hand:

$$y_{nt} = x'_{nt}\beta + \gamma_n + \varepsilon_{nt}, \tag{2.2}$$

waarin γ_n het individuele effect van staat n aangeeft. Omdat het gemiddelde van γ_n precies dezelfde rol speelt als de constante term β_1, is hun beider effect niet te ontwarren en wordt gewoonlijk, zonder verlies van algemeenheid, het gemiddelde van de γ_n gelijk aan nul gesteld.

Bekende voorbeelden uit de econometrische literatuur van een dergelijk model zijn het artikel van Mundlak (1961), waarin de produktiviteit (y) van een aantal landbouwbedrijven over de tijd wordt gevolgd (γ_n kan dan worden geïnterpreteerd als de – niet direct meetbare – kwaliteit van de leiding van het bedrijf n), en de studie van bijvoorbeeld Griliches (1977) over de 'returns to schooling'; y_{nt} is dan het inkomen van individu n op tijdstip t, en γ_n zijn of haar 'ability'.

Model (2.2) laat een tweede voordeel zien van het werken met panelgegevens, namelijk het rekening houden met een niet direct waarneembare variabele (kwaliteit van de leiding, 'ability'), die verschilt over de waarnemingseenheden maar die *sui generis* constant is over de tijd.

Naar analogie met het opnemen van γ_n wordt dikwijls ook een tijdseffect in het model opgenomen, wat tot een nog algemenere specificatie leidt:

$$y_{nt} = x'_{nt}\beta + \gamma_n + \delta_t + \varepsilon_{nt}, \tag{2.3}$$

waarbij δ_t het specifieke effect van jaar t voorstelt, weer met gemiddelde nul. Bij het eerdere voorbeeld van de produktiviteit van de landbouwbedrijven kan men bij δ_t denken aan het gezamenlijke effect van de niet expliciet opgenomen, jaarspecifieke factoren als het weer en de conjunctuur en iedere andere voor alle bedrijven gelijkelijk werkende invloed. Het opnemen van δ_t naast γ_n compliceert de analyse uiteraard enigszins en omdat in de meeste praktijkgevallen de ontwikkeling van jaar tot jaar geleidelijk is maar de variatie over de waarnemingseenheden naar verhouding groot, wordt δ_t veelal weggelaten zonder dat dat een ernstige specificatiefout oplevert.

3. SCHATTINGSASPECTEN

Hoe moet model (2.3) worden geschat? Dat hangt af van de wijze waarop de γ_n en de δ_t worden opgevat, *vast* of *stochastisch*. In het vaste-effectenmodel (VEM) worden de γ_n en de δ_t opgevat als vaste, te schatten parameters, en in het stochastische-effectenmodel (SEM), ook wel het variantie-componentenmodel geheten, als onafhankelijke trekkingen uit twee kansverdelingen, ieder met verwachting nul, een die de γ_n genereert, en een de δ_t. In dat geval zijn alleen de varianties van die kansverdelingen (σ_γ^2 en σ_δ^2) de te schatten parameters. Anders gezegd, in het VEM bestaat de constante term uit drie componenten ($\beta_1+\gamma_n+\delta_t$ in plaats van β_1), en in het SEM de storingsterm; in plaats van uit ε_{nt} zoals in (2.1) bestaat hij nu uit $\gamma_n+\delta_t+\varepsilon_{nt}$.

Het schatten van het VEM is een eenvoudige zaak; omdat de γ_n en de δ_t vaste parameters zijn, is het VEM niets anders dan een gewoon regressiemodel waaraan $N+T$ dummy-variabelen zijn toegevoegd, corresponderend met de regressiecoëfficiënten $\gamma_1, \ldots, \gamma_N$ en $\delta_1, \ldots, \delta_T$. De enige complicaties worden veroorzaakt door de restricties dat de γ_n en de δ_t ieder tot nul optellen, en dat $N+T$ een groot getal kan zijn zodat de computer het niet meer aan kan. Beide problemen kunnen worden opgelost door, in plaats van model (2.3), model (2.1) te schatten nadat de data de zogenaamde *covariantietransformatie* hebben ondergaan, waarbij y_{nt} vervangen wordt door $y_{nt}-y_{n.}-y_{.t}+y_{..}$ (een punt op de plaats van een index geeft het gemiddelde over die index aan), en ieder van de elementen van x_{nt} op analoge wijze is bewerkt; de constante term wordt weggelaten. Het is eenvoudig aan te tonen dat beide methoden, (2.3) rechtstreeks schatten of (2.1) na de covariantietransformatie, dezelfde schatter $\hat{\beta}$ voor β oplevert. (In het eenvoudiger model (2.2) volstaat de transformatie van y_{nt} tot $y_{nt}-y_{n.}$, en analoog voor de x_{nt}.) Schatters voor de γ_n en de δ_t worden vervolgens gevonden door de residuen $y_{nt}-x_{nt}'\hat{\beta}$ te middelen over respectievelijk de tijd en de waarnemingseenheden.

Lastiger is het schatten van het SEM. Zoals gezegd is het SEM een regressiemodel met een samengestelde storingsterm. Noem die u_{nt}, dan geldt

$$u_{nt}=\gamma_n+\delta_t+\varepsilon_{nt}, \tag{3.1}$$

zodat

$$\text{covariantie}\,(u_{nt}, u_{ms}) = \sigma_\varepsilon^2 \begin{cases} 1+\sigma_\delta^2/\sigma_\varepsilon^2+\sigma_\delta^2/\sigma_\varepsilon^2 & m=n,\ s=t \\ \sigma_\gamma^2/\sigma_\varepsilon^2 & m=n,\ s\neq t \\ \sigma_\delta^2/\sigma_\varepsilon^2 & m\neq n,\ s=t \\ 0 & m\neq n,\ s\neq t \end{cases} \tag{3.2}$$

Ter vergelijking: bij gewone regressie is die covariantie gelijk aan σ_ε^2 (als $m=n$ en $s=t$) of 0. Er treedt dus heteroskedasticiteit en afhankelijkheid op, en de juiste schattingsmethode is die van de gegeneraliseerde kleinste kwadraten (GKK). Om die te kunnen toepassen moet de covariantie tussen ieder van de storingen bekend zijn (althans op een scalaire factor na), en dat is natuurlijk niet het geval: afgezien van de factor σ_ε^2 komen in (3.2) de onbekende verhoudingen $\sigma_\gamma^2/\sigma_\varepsilon^2$ en $\sigma_\delta^2/\sigma_\varepsilon^2$ voor. Een twee-staps-procedure is dan nodig: eerst moe-

ten voor deze grootheden (asymptotisch rake) schatters worden gevonden, en vervolgens kan, gebruik makend van de gevonden waarden, in de tweede ronde de GKK-schatter voor β worden berekend. In de literatuur over dit onderwerp is nogal wat aandacht besteed aan de vraag hoe de variantiecomponenten σ_γ^2 en σ_δ^2 moeten worden berekend; zo onderscheiden Maddala en Mount (1973) elf methoden, en onderwerpt Baltagi (1981) negen methoden aan een vergelijkende studie. In wezen maakt de keuze niet zoveel uit waar het de schatting van β in de tweede ronde betreft. Een goede en eenvoudige methode is het VEM te schatten en de daaruit berekenbare variantie van de schattingen van de γ_n te gebruiken als schatting voor σ_γ^2, en die van de δ_n voor σ_δ^2 (zie bijvoorbeeld Amemiya, 1971, Mazodier, 1972).

Uit het geval $m=n$, $s \neq t$, in (3.2) blijkt dat het SEM impliceert dat de storingen in dit model, voor een gegeven waarnemingseenheid maar voor verschillende tijdstippen, eenzelfde covariantie hebben. Hoever die tijdstippen uit elkaar liggen speelt geen rol. Naast het eerder genoemde bezwaar tegen model (2.3) versus (2.2) (gecompliceerder zonder dat dat veel oplevert, want $\sigma_\gamma^2 \gg \sigma_\delta^2$ in de praktijk) is dit een tweede bezwaar: afhankelijkheden over de tijd zijn bevredigender te modelleren dan (impliciet) is gebeurd door het invoeren van het tijdseffect δ_t.

4. SEM OF VEM?

Gegeven deze bespreking van de schattingsaspecten van SEM en VEM resteert de vraag: welk model moet een onderzoeker nu kiezen? Deze vraag, die reeds in de variantie-analyse de nodige aandacht heeft gekregen (Scheffé, 1959, Searle, 1971), is ook in de econometrische literatuur uitgebreid besproken. De voordelen van het VEM zijn duidelijk: het is eenvoudig te schatten, en een onderzoeker die het SEM wil schatten zal – zie boven – er goed aan doen eerst het VEM te schatten om aan schattingen voor σ_γ^2 en σ_δ^2 te komen. Bovendien wordt bij het SEM verondersteld dat de γ_n en δ_t onafhankelijke trekkingen uit een kansverdeling zijn – dat wil zeggen, ook onafhankelijk van x_{nt}. Als aan die veronderstelling niet is voldaan, doet zich dus het fenomeen voor van afhankelijkheid tussen de verklarende variabele in een regressie en de storingsterm. Dit leidt, zoals in ieder econometrieleerboek is beschreven, tot schatters voor β die asymptotisch niet-raak zijn (dat wil zeggen, niet naar de ware waarden toegaan als het aantal waarnemingen naar oneindig toegaat) en daarom niet voldoen aan de zwakste en tegelijk meest essentiële eis die aan een schatter gesteld kan worden.

De veronderstelling van onafhankelijkheid is in de praktijk vaak dubieus. Zo is het, bij het voorbeeld van de landbouwbedrijven, maar de vraag of de kwaliteit van de leiding van een bedrijf ongecorreleerd is met *alle* verklarende variabelen! Het is duidelijk waarom dit afhankelijkheidsprobleem bij het VEM niet bestaat: bij het berekenen van $\hat{\beta}$ wordt de covariantietransformatie toegepast, waarmee de γ_n en δ_t worden geëlimineerd, en daarmee ook ieder probleem dat zij kunnen veroorzaken.

Dit alles suggereert uniforme superioriteit van VEM. Dat is echter niet zo. Om dit in te zien is het het gemakkelijkst het eenvoudige model (2.2) te bezien. Bij het schatten hiervan onder de VEM-specificatie gaat het om transforma-

ties als

$$\tilde{y}_{nt} = y_{nt} - y_{n.}, \tag{4.1}$$

zodat uiteraard

$$\sum_{t=1}^{T} \tilde{y}_{nt} = 0, \tag{4.2}$$

voor $n = 1, \ldots, N$. Er zijn dus N restricties op de waarnemingen geïntroduceerd, oftewel, er zijn in zekere zin N van de in totaal NT waarnemingen verdwenen. Het enige dat van y_{nt} meedoet is het verschil met $y_{n.}$, het tijdsgemiddelde; dat gemiddelde zelf speelt geen rol. De VEM-schatter verwaarloost dus bij het schatten van β alle kennis die uit het rechtstreeks vergelijken van individuen kan worden afgeleid. Onder de VEM-specificatie werkt men, kortom, met minder informatie dan oorspronkelijk beschikbaar – een deel daarvan wordt namelijk gebruikt voor het schatten van de γ_n – en de prijs hiervoor is een geringere nauwkeurigheid van de schatter $\hat{\beta}$.

Men kan vrij eenvoudig afleiden dat het schatten van (2.2) onder de SEM-specificatie ook kan worden geïnterpreteerd als het schatten van (2.1) na een passende transformatie. Deze luidt:

$$\tilde{\tilde{y}}_{nt} \equiv y_{nt} - \phi y_{n.} = \tilde{y}_{nt} + (1 - \phi) y_{n.} \tag{4.3}$$

(en weer analoog voor de x_{nt}), met ϕ een grootheid die van $\sigma_\gamma^2/\sigma_\varepsilon^2$ afhangt:

$$\phi \equiv 1 - (1 + T\sigma_\gamma^2/\sigma_\varepsilon^2)^{-1/2}. \tag{4.4}$$

Uit (4.3) blijkt dat de SEM-schatter aan de tijdgemiddelden $y_{n.}$ wel een zelfstandige rol toekent (naast het ontdoen van y_{nt} van zijn tijdgemiddelde, zie (4.1)) en daarom in het algemeen nauwkeuriger zal zijn.

Daarbij moet wel worden opgemerkt dat het hier om een voordeel gaat dat zich in principe alleen voordoet bij weinig waarnemingen in de tijd; immers, het aantal onafhankelijke waarnemingen bij VEM (versie (2.2)) is $(T-1)/T$ maal dat bij SEM, en als T maar groot genoeg wordt worden beide modellen equivalent. Nu zal bij veel panels T niet al te groot zijn, en lijkt SEM in het voordeel. Het is niet goed mogelijk om de nauwkeurigheid van $\hat{\beta}$ onder VEM ten opzichte van die onder SEM langs wiskundige weg te vergelijken, omdat de laatste een moeilijk te kwantificeren onzekerheidsfactor bevat als gevolg van het feit dat voor $\sigma_\gamma^2/\sigma_\varepsilon^2$ een schatting moet worden gebruikt, wat nauwkeurigheid kost.

Dit alles suggereert dat er geen a priori reden is om SEM of VEM te verkiezen: de onderzoeker zou beiden kunnen hanteren, en kiezen wat hem het beste bevalt, zolang hij tenminste geen inhoudelijke keuzegrond heeft: het kan voor de hand liggen om bij sommige panels VEM te kiezen (een onderzoeker kan zeer wel geïnteresseerd zijn in de specifieke effecten van Groningen, Friesland, enz.), en om bij andere panels SEM te kiezen – de specifieke effecten van een paar duizend aselect getrokken geënquêteerden zullen over het algemeen niet erg interessant zijn. Erg hard is dit allemaal niet, en het mag dan ook geen verwondering wekken dat het verschil tussen beide schatters in de literatuur de nodige aandacht heeft gekregen. Deze aandacht richtte zich op de zwakke plek

in SEM, namelijk de veronderstelde onafhankelijkheid van de γ_n en de x_{nt}.

In een opmerkelijk artikel vraagt Mundlak (1978) zich af waarom er überhaupt twee schatters bestaan voor in wezen hetzelfde model, en laat vervolgens zien dat er maar één schatter is. Voor deze verrassende conclusie is het nodig de onafhankelijkheid van de γ_n en de x_{nt} te vervangen door een expliciete afhankelijkheid:

$$\gamma_n = x'_{n.}\pi + \xi_n, \tag{4.5}$$

met $x_{n.}$ de vector van gemiddelden (over de tijd) van de elementen van x_{nt}, π een coëfficiëntenvector, en ξ_n een storingsterm, met verwachting nul en onafhankelijk van x. Als men (2.2) met (4.5) uitbreidt, is het een kwestie van enige algebra om te laten zien dat $\hat{\beta}$ onder SEM gelijk is aan $\hat{\beta}$ onder VEM. Bovendien kan ook π eenvoudig worden geschat, en kan de hypothese $\pi = 0$ worden getoetst; als die hypothese niet behoeft te worden verworpen was de oorspronkelijke onafhankelijkheidsveronderstelling toch houdbaar, en mag men de SEM-schatter gebruiken.

Een andere toets op de onafhankelijkheid, die niet gebaseerd is op het verband (4.5) tussen het specifieke effect en het gemiddelde van een aantal onafhankelijke variabelen, wordt gegeven door Hausman (1978). De (asymptotische) gelijkheid van de VEM- en de SEM-schatters betekent dat het verschil tussen beide een indicatie over de juistheid van de onafhankelijkheidsveronderstelling oplevert. Hausman laat zien hoe uit dit verschil een eenvoudige toets is af te leiden.

Is er eenmaal afhankelijkheid geconstateerd, dan lijkt VEM het aangewezen model, en dat is het in de meeste gevallen ook. Er zijn echter twee redenen die tegen VEM kunnen pleiten. De eerste is de al eerder genoemde onnauwkeurigheid als gevolg van de beperkte rol die de tijdgemiddelden spelen (zie (4.1)). De tweede reden is de mogelijke aanwezigheid van verklarende variabelen die over de tijd niet veranderen; als die variabelen worden getransformeerd volgens (4.1), worden ze identiek nul. Datzelfde gebeurt met de effecten, en dat was daarbij precies de bedoeling; maar het elimineren van tijd-invariante variabelen kan een verlies zijn omdat de onderzoeker juist in hun invloed is geïnteresseerd.

Een voorbeeld van zulk onderzoek wordt geleverd door een variant van de eerder genoemde studies naar de invloed van schoolopleiding op inkomen. Deze luidt:

$$y_{nt} = x'_{nt}\beta + z_n\alpha + \gamma_n + \varepsilon_{nt}, \tag{4.6}$$

waarin y_{nt} weer het inkomen van individu n in het jaar t aangeeft, en z_n het aantal jaren schoolopleiding; de coëfficiënt α geeft dus de bijdrage aan, aan het jaarinkomen, van een jaar scholing, en is de meest interessante grootheid in dit model. Nu zijn γ_n en z_n waarschijnlijk gecorreleerd (γ_n was eerder al geïnterpreteerd als 'ability', en als je daar veel van hebt zul je ook wel een langere opleiding volgen), dus SEM is onbruikbaar, maar VEM ook: dan zou z_n verdwijnen. Dit probleem is door Griliches (1977) opgelost door middel van een zogenaamde instrumentele variabele: stel dat van ieder individu nog een variabele (zeg: h_n) bekend is die *niet* met γ_n en ε_{nt} correleert maar wel met z_n, dan volgt uit (4.5), als we eenvoudigheidshalve het stuk $x'_{nt}\beta$ weglaten en aan-

nemen dat alle y_n en z_n in afwijking van hun totale gemiddelden zijn gemeten,

$$y_{n.} = z_n\alpha + \gamma_n + \varepsilon_{n.} \tag{4.7}$$

en dus

$$\frac{1}{N}\sum_{n=1}^{N} h_n y_{n.} = \left(\frac{1}{N}\sum_{n=1}^{N} h_n z_n\right)\alpha + \frac{1}{N}\sum_{n=1}^{N} h_n(\gamma_n + \varepsilon_{n.}) \tag{4.8}$$

oftewel, als we steekproefcorrelaties met r aangeven:

$$r_{hy} = r_{hz}\alpha + r_{h\gamma} + r_{h\varepsilon} \tag{4.9}$$

in voor de hand liggende notatie. De laatste twee correlaties zijn volgens veronderstelling gelijk aan 0, zodat het gebruik van het instrument h het mogelijk maakt een zuivere en asymptotisch rake schatter voor α af te leiden: $\hat{\alpha} = r_{hy}/r_{hz}$.

Als instrument gebruikt Griliches (1977) een variabele die betrekking heeft op de achtergronden van het gezin van individu n. Het is daarbij de vraag of deze variabele inderdaad *niet* met γ_n correleert. Een verwante oplossing voor het probleem is recentelijk aangedragen door Hausman en Taylor (1981). Zij beschouwen die elementen x_{0nt} van x_{nt} die *niet* met γ_n correleren (zo die er zijn, anders heeft het geen zin), en gebruiken die op twee manieren: $x_{0nt} - x_{0n.}$ wordt op de gewone manier (VEM) gebruikt om β te schatten, en de bij VEM verwaarloosde $x_{0n.}$ wordt gebruikt als instrument om α te schatten: aan de eis van onafhankelijkheid van x_{0nt} en γ_n (en dus ook van $x_{0n.}$ en γ_n) was immers tevoren per assumptie voldaan. De structuur van de data levert dus een instrument op 'natuurlijke' wijze. Het artikel van Hausman en Taylor, dat verschillende onafhankelijkheidstoetsen bespreekt en een aantal asymptotisch meest nauwkeurige schattingsmethoden presenteert, laat zien hoe het SEM een aantal problemen waar het VEM op stukloopt, alsnog kan oplossen.

5. VARIANTEN EN ANDERE MODELLEN

Het voorgaande mag de indruk hebben gewekt dat de econometrie van de paneldata identiek is met de analyse van het VEM of het SEM. Dit is allerminst het geval, hoewel genoemde modellen vanwege hun eenvoud wel de meeste toepassing hebben gevonden, en bespreking ervan een aantal problemen illustreert waarmee de econometrist zich bezig houdt. In deze paragraaf worden een aantal varianten kort besproken, en enige andere modellen getoond.

Het VEM is een model waarin de constante term (β_1) in drieën wordt opgesplitst. Een voor de hand liggende uitbreiding is om *alle* elementen van β, dus ook de 'echte' regressiecoëfficiënten, op dergelijke wijze te differentiëren. Aldus verkrijgt men het zogenaamde Variabele-Coëfficiënten-Model (VCM). Logisch gezien is de stap van VEM naar VCM een kleine, maar conceptueel is er nogal wat verschil: γ_n representeert het niveau-effect van alle niet expliciet opgenomen, voor n specifieke invloeden, terwijl bij het VCM wordt verondersteld dat eenheid n op een eigen wijze reageert op de regressoren.

Bij het VCM worden weer een vaste en een stochastische versie onderscheiden; zie bijvoorbeeld Malinvaud en Milleron (1970), Swamy (1971), Hsiao

(1974, 1975), waarin zowel een twee- als drie-componentenmodel (analoog aan (2.2) en (2.3)) wordt behandeld. Een variant hierop is het door Wansbeek en Kapteyn (1978, 1982) ontwikkelde model

$$y_{nt} = x'_{nt}\beta + z'_{1n}\alpha_{1t} + z'_{2t}\alpha_{2n} + \varepsilon_{nt}, \tag{5.1}$$

met z_{1n} en z_{2t} vectoren van, respectievelijk, tijd-invariante en tijd-afhankelijke variabelen, en α_{1t} en α_{2n} vectoren van regressiecoëfficiënten. Model (5.1) is een generalisering van (4.6) en (2.3) en van allerlei andere modellen voor het analyseren van variabelen met dubbele index. Zie Wansbeek (1980) voor een uitgebreide bespreking van de schattingstechnische aspecten van dit model.

Tot nu toe is het aspect 'tijd' behandeld op een manier die volkomen analoog is aan het aspect 'waarnemingseenheid'. Er is natuurlijk een essentieel verschil en dat is, dat processen in de tijd gewoonlijk met een zekere geleidelijkheid plaatsvinden: de waarde van een variabele op tijdstip t wordt mede bepaald door zijn waarde op $t-1$. Dan wordt (2.3):

$$y_{nt} = x'_{nt}\beta + y_{n,t-1}\alpha + \gamma_n + \delta_t + \varepsilon_{nt}, \tag{5.2}$$

dus (2.2) met de vertraagde endogene variabele (y) als extra regressor opgenomen. De coëfficiënt α, $0 \leqslant \alpha < 1$, geeft aan in welke mate de vorige waarde van y de huidige waarde bepaalt. De interpretatie van de coëfficiëntenvector β van de 'gewone' regressor x behoeft daardoor wat verfijning: β geeft de invloed van x aan *naast* die van de vertraagde endogeen; in een toestand van evenwicht ($y_{nt} = y_{n,t-1} \equiv y_{n*}$) volgt, door invulling hiervan in (5.2), dat de invloed van x gelijk is aan $\beta/(1-\alpha)$:

$$y_{n*} = x_{nt}\beta + y_{n*}\alpha + \gamma_n + \delta_t + \varepsilon_{nt} \tag{5.3}$$

oftewel

$$y_{n*} = x_{nt}\beta/(1-\alpha) + (\gamma_n + \delta_t + \varepsilon_{nt})/(1-\alpha). \tag{5.4}$$

Hierom wordt $\beta/(1-\alpha)$ wel de vector van *lange-termijncoëfficiënten* genoemd, en β die van de *korte-termijncoëfficiënten*.

Uit de econometrische literatuur (bijvoorbeeld Theil, 1971) blijkt dat modellen met vertraagde endogenen lastiger te hanteren zijn dan gewone modellen, en dit geldt a fortiori voor de wat meer gecompliceerde modellen zoals hier beschouwd. Het voert te ver om deze problemen hier uit te werken. Er zij volstaan met de constatering dat ook hier een VEM- en een SEM-versie kunnen worden onderscheiden, en dat het voor beide versies niet-triviaal is om nauwkeurige en asymptotisch rake schatters af te leiden: bij het VEM omdat uit simulatiestudies (Maddala, 1971) en uit latere analytische resultaten (Nickell, 1981) blijkt dat de gewone schattingsmethode voor kleine T in hoge mate onzuivere schatters oplevert (de onzuiverheid verdwijnt bij grote T, maar dat is in de praktijk een schrale troost), en bij het SEM omdat dit een model is waar tegelijkertijd de vertraagde endogeen en afhankelijkheid tussen storingstermen optreden, wat reeds in een gewoon regressiemodel (dat wil zeggen (5.2) voor $N=1$) de analyse compliceert, zoals de gewone econometrische literatuur ons leert.

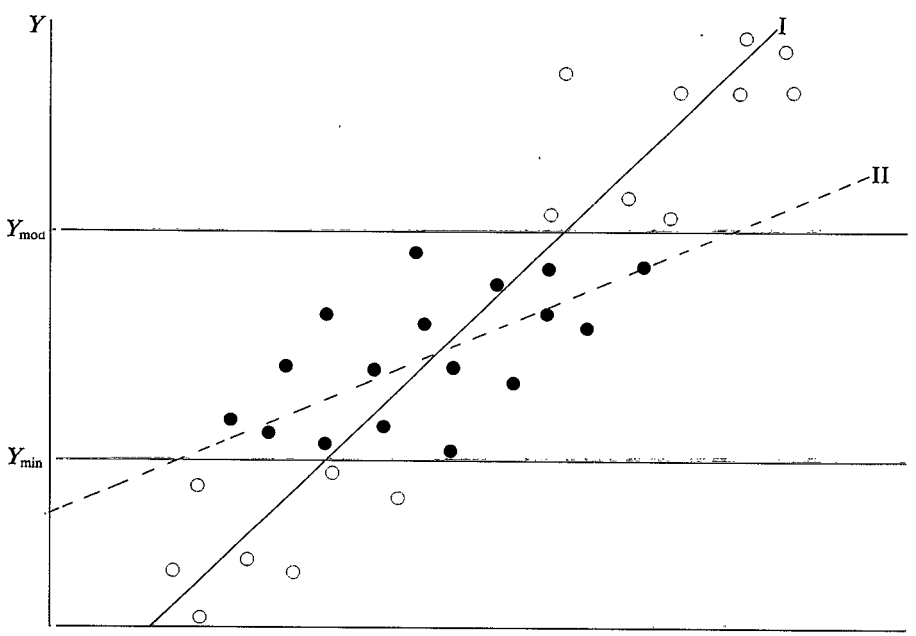
6. ONVOLLEDIGE PANELBESTANDEN

Lag de nadruk tot nu toe, in onze bespreking en in de geciteerde artikelen, op de aanwezigheid van een volledig panelbestand, in deze paragraaf gaan we in op de specifieke problemen die zich bij een onvolledig panelbestand voordoen. Zoals in de inleiding aangeduid, komen dergelijke bestanden vooral voor als men te doen heeft met een steekproef van personen of gezinnen die herhaaldelijk worden geënquêteerd: er ontstaat uitval.

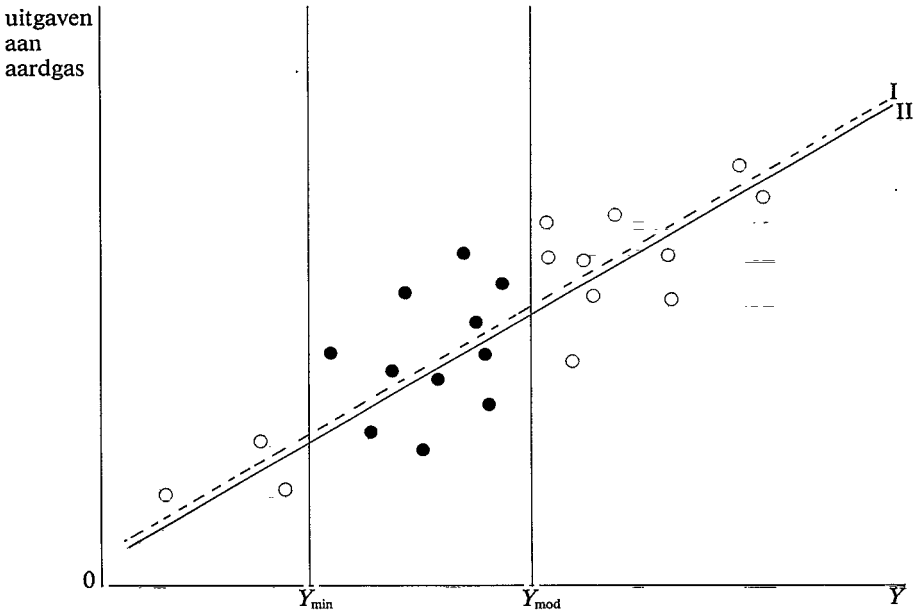
Er zijn twee soorten uitval met ieder hun eigen consequenties, en wel *selectieve* en *aselecte* uitval. Bij selectieve uitval hangt het wel of niet uitvallen samen met de storingsterm in het model dat men wil schatten; bij aselecte uitval is die samenhang er niet. Een voorbeeld moge dit onderscheid verduidelijken en laat zien dat selectieve uitval (of, in het algemeen, selectieve non-respons; de kwestie speelt evenzeer bij een enkele steekproef) een groter probleem vormt dan aselecte non-respons.

Bezie figuur 1A. Een onderzoeker wil, geïnspireerd door de discussie in paragraaf 4, ook voor Nederland de invloed van schoolopleiding op inkomen onderzoeken. Daartoe trekt hij een steekproef, stelt de juiste vragen, krijgt een puntenwolk (zwarte en witte stippen) en vindt, als zijn steekproef maar groot genoeg is en zijn model goed gespecificeerd, het ware verband, lijn I. Een andere onderzoeker wil hetzelfde bestand analyseren, maar vindt alleen de groep tussen minimum en modaal relevant; alle andere waarnemingen worden verwijderd. Wat resulteert is lijn II, en hoe groot de steekproef ook is, de helling van lijn I wordt systematisch onderschat.

Figuur 1B beschrijft de omgekeerde situatie. Een onderzoeker wil de invloed van inkomen op de uitgaven aan aardgas weten, vindt ook dat alleen de



Figuur 1A. Selectie op een endogeen aantal schooljaren



Figuur 1B. Aselecte non-respons

groep tussen minimum en modaal meetelt, selecteert daarop, en krijgt, omdat hij op de *exogene* variabele selecteert, wèl een asymptotisch rake schatter van het verband tussen inkomen en uitgaven aan aardgas. (Daarmee is niet gezegd dat het dan ook een *slimme* aanpak is: als je de te enquêteren personen van tevoren op inkomen kunt selecteren en je wilt een verband schatten, met inkomen als exogeen, dat voor alle inkomensgroepen hetzelfde is, dan is het optimaal om zoveel mogelijk spreiding in de inkomens te hebben om een zo nauwkeurig mogelijke schatting te krijgen; zie bijvoorbeeld Aigner en Morris, 1979.)

De moraal is dat selectie op de afhankelijke variabele in principe leidt tot *niet*-rake schatters. Veelal zullen dan de regressiecoëfficiënten worden onderschat. Omdat de afhankelijke variabele bestaat uit een systematisch stuk en de storingsterm, is een andere en iets algemenere formulering van dit probleem: selectie die samenhangt met de storingsterm.

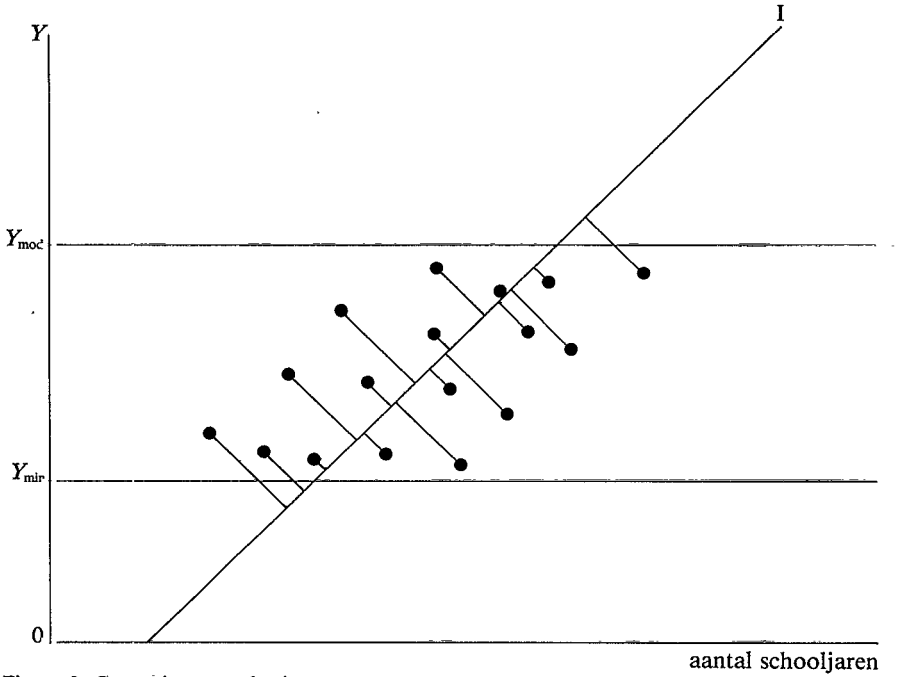
Bij aselecte non-respons doet het probleem van de niet-rake schatters zich niet voor. Bij een panel met aselecte uitval kan de onderzoeker toch wel voor wat problemen komen te staan, maar die zijn primair rekentechnisch van aard. Het probleem is dat van de 'unbalanced data' in de variantie-analyse (zie bijvoorbeeld Searle, 1971), en dit houdt in dat de eenvoudige, in paragraaf 3 besproken transformaties niet meer zonder meer kunnen worden toegepast.

De onderzoeker kan twee dingen doen. Hij kan, als er niet al te veel (aselecte) uitval is, alle onvolledige respondenten uit zijn dataset verwijderen en met een volledig bestand doorgaan. Griliches e.a. (1978) laten zien dat dit zelfs bij een vrij groot aantal onvolledige waarnemingen een verstandige aanpak is. Als er dusdoende echter te weinig (of misschien wel helemaal geen – denk aan roterende panels waar bijvoorbeeld ieder jaar de helft wordt ververst) respon-

denten overblijven, dan moet een andere aanpak worden gekozen. Deze kan bestaan uit het toepassen van een, rekentechnisch onappetijtelijke, methode uit de 'unbalanced data'-theorie (zie wederom Searle, 1971), of, bij SEM, uit het toepassen van de methode van de grootste aannemelijkheid om β, σ_γ^2, σ_δ^2 en σ_ε^2 tegelijk te schatten. Bij volledige waarnemingen is deze methode uiteraard ook beschikbaar (Amemiya, 1971) maar door zijn iteratieve karakter niet aanbevelenswaardig, gegeven het bestaan van veel eenvoudiger, equivalente methoden (tenzij men het eenvoudige model (2.2) wil schatten en over een programma als LISREL beschikt); bij onvolledige waarnemingen verdwijnt dit nadeel en lijkt toepassing van de grootste-aannemelijkheidsmethode zinvol. Veel ervaring is er overigens nog niet. Wansbeek en Kapteyn (1981) leiden de relaties af waaraan de meest aannemelijke schatters in het SEM bij onvolledige waarnemingen moeten voldoen, en Biørn (1981) past de methode toe op een panel afgeleid uit het Noorse budgetonderzoek waarin 209 respondenten in 1975 en 1976 meededen, en 209 anderen in 1976 en 1977. Door deze simpele structuur wordt het berekenen van de meest aannemelijke schatters vereenvoudigd.

Bezien wij nu het probleem dat optreedt bij selectieve uitval. Zoals aangegeven leidt verwaarlozing van dit probleem tot een systematische vertekening van de regressiecoëfficiënten. Wil men tot juiste resultaten komen, dan is informatie nodig over het proces dat de uitval genereert, en dit is de lijn die de laatste jaren in de econometrische literatuur is gevolgd (bijvoorbeeld Heckman, 1976, 1979): het te schatten model wordt uitgebreid met een relatie die aangeeft welke factoren het wel of niet responderen bepalen. Een dergelijke responsvergelijking ziet er als volgt uit. De te verklaren variabele is dichotoom: wel of geen respons, en de verklarende variabelen zijn gewoonlijk een aantal sociaal-economische kenmerken van de respondenten en non-respondenten, zoals inkomen, het wel of niet in een grote stad wonen, en leeftijd. (Zie bijvoorbeeld Van de Stadt en Wansbeek (1982) voor een toepassing.) Ook heeft de responsvergelijking een storingsterm, en wanneer die correleert met die van de oorspronkelijke relatie is er selectieve non-respons. Er zijn methoden ontwikkeld om de covariantie tussen die storingen te schatten, waarmee de aanwezigheid van selectiviteit, zo die niet meteen duidelijk is zoals in figuur 1A, kan worden getoetst.

Een methode om voor selectiviteit bij uitval in panels te corrigeren is gegeven door Hausman en Wise (1979). Dit artikel, dat in 1980 bekroond is als de beste in 'Econometrica' gepubliceerde empirische studie in de voorafgaande vijf jaar, laat zien hoe model (2.2) kan worden uitgebreid met een responsvergelijking en hoe optimale schatters er uitzien. Een interpretatie van het gebruik dat wordt gemaakt van de extra-informatie, afkomstig van de responsvergelijking, is in figuur 2 aangegeven. De waarnemingen (zwarte stippen) in deze figuur zijn dezelfde als in figuur 1A. Uit die figuur bleek dat gewone regressie tot een onjuiste lijn (lijn II) leidde. Bij gewone regressie wordt de som van de kwadraten van de afstanden tussen de punten en de regressielijn in *verticale* zin geminimaliseerd. Door de afstanden niet verticaal, maar onder een bepaalde hoek te meten, is het mogelijk op grond van een selectieve steekproef de juiste regressielijn (lijn I) te vinden; zie figuur 2. De extra informatie uit de responsvergelijking geeft nu aan hoe groot deze hoek moet worden gekozen.



Figuur 2. Correctie voor selectieve non-respons

Dat het niet gaat om academische spielerei maar om belangrijke effecten blijkt uit een eerder artikel van Hausman en Wise (1976). Zij her-analyseerden de uitkomsten van het 'New Jersey Negative Income Tax Experiment', een miljoenen dollars duur sociaal experiment waarin de reactie, vooral ten aanzien van arbeidsmarktgedrag, van individuen op alternatieve sociale-zekerheidsmaatregelen wordt nagegaan. Rekening houdend met selectiviteit vonden zij aanmerkelijk grotere effecten van inkomensmaatregelen op het arbeidsaanbod dan in eerder onderzoek. Verwaarlozing van selectiviteit kan tot verkeerde beleidsbeslissingen leiden.

7. CONCLUSIE

In dit artikel is getracht aan te geven hoe de econometrie met panels omgaat. De behandeling was niet-technisch en weinig diepgravend, en verwaarloosde een aantal interessante aspecten; zo is bijvoorbeeld niet gesproken over het voorspellen met panels (Hester, 1978, Wansbeek en Kapteyn, 1978, Taub, 1979), rekentechnische aspecten (Wansbeek, 1980) of het werken met kwalitatieve data. Dit laatste aspect is door Chamberlain (1980) behandeld voor het geval waar y_{nt} alleen de waarden 0 of 1 aanneemt. Er blijkt dan een verband te kunnen worden gelegd tussen het SEM en het zogenaamde log-lineaire model, een veelgebruikte methode om frequentietabellen te analyseren (zie bijvoorbeeld Fienberg, 1977). Hopelijk heeft dit artikel duidelijk gemaakt dat het analyseren van panel-gegevens in de econometrie in volle ontwikkeling is.

REFERENTIES

Aigner, D.J. en C.N. Morris (red.), 1979. *Experimental design in econometrics.* Annals of Applied Econometrics 1979-2.

Amemiya, T., 1971. *The estimation of the variances in a variance-components model.* International Economic Review 12, pp. 1-13.

Balestra, P. and M. Nerlove, 1966. *Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas.* Econometrica 34, pp. 585-612.

Baltagi, B.H., 1981. *Pooling: an experimental study of alternative testing and estimation procedures in a two-way error component model.* Journal of Econometrics 17, pp. 21-49.

Biørn, E., 1981. *Estimating economic relations from incomplete cross-section/time-series data.* Journal of Econometrics 16, pp. 221-236.

Chamberlain, G., 1980, *Analysis of covariance with qualitative data.* Review of Economic Studies 47, pp. 225-238.

Fienberg, S.E., 1977. *The analysis of cross-classified categorical data.* MIT Press, Cambridge, Massachusetts.

Griliches, Z., 1977. *Estimating the returns to schooling: some econometric problems.* Econometrica 45, pp. 1-22.

Griliches, Z., B.H. Hall en J.A. Hausman, 1978. *Missing data and self-selection in large panels.* Annales de l'INSEE 30-31, pp. 137-176.

Hausman, J.A., 1978. *Specification tests in econometrics.* Econometrica 46, pp. 1251-1271.

Hausman, J.A. en W.E. Taylor, 1981. *Panel data and unobservable individual effects.* Econometrica 49, pp. 1377-1398.

Hausman, J.A. en D.A. Wise, 1976. *The evaluation of results from truncated samples: The New Jersey Negative Income Tax Experiment.* Annals of Economic and Social Measurement 5, pp. 421-445.

Hausman, J.A. en D.A. Wise, 1979. *Attrition bias in experimental and panel data: The Gary Income Maintenance Experiment.* Econometrica 47, pp. 455-473.

Heckman, J.J. 1976. *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models.* Annals of Economic and Social Measurement 5, pp. 475-492.

Heckman, J.J., 1979. *Sample selection bias as a specification error.* Econometrica 47, pp. 153-161.

Hester, D.H., 1978. *Dynamic portfolio behavior and market clearing by weekly reporting banks.* Annales de l'INSEE 30-31, pp. 297-330.

Hsiao, C., 1974. *Statistical inference for a model with both random cross-sectional and time effects.* International Economic Review 15, pp. 12-30.

Hsiao, C., 1975. *Some estimation methods for a random coefficients model.* Econometrica 43, pp. 305-326.

Judge, G.G., W.E. Griffiths, R.C. Hill en T.-C. Lee, 1980. *The theory and practice of econometrics,* Wiley, New York.

Kapteyn, A., 1977. *Consumentengedrag en inkomenswaardering.* In: J.W. van der Ben (red.) Marktonderzoek en consumentengedrag, Stenfert Kroese, Leiden.

Kapteyn, A. en T.J. Wansbeek, 1978. *Beter meten en voorspellen met behulp van panel-gegevens.* Economisch-Statistische Berichten 63, pp. 43-47.

Kapteyn, A., en T.J. Wansbeek, 1982. *The individual welfare function: measurement, explanation and policy applications.* Statistical Studies 32, Staatsuitgeverij, 's-Gravenhage.

Kmenta, J., 1971. *Elements of econometrics,* Macmillan, New York.

Maddala, G.S., 1971. *The use of variance components models in pooling cross-section and time series data,* Econometrica 39, pp. 341-358.

Maddala, G.S., 1977. *Econometrics,* McGraw-Hill, New York.

Maddala, G.S. en T.D. Mount, 1973. *A comparative study of alternative estimators for variance components models used in econometric applications.* Journal of the American Statistical Association 68, pp. 324-328.

Malinvaud, E. en J.C. Milleron, 1970. *Généralisation des modèles d'analyse de la variance dans le cas de séries temporelles de coupes instantanées; Conditions d'identifiabilité et méthodes d'estimation,* INSEE rapport 348/930, Parijs.

Mazodier, P., 1972. *l'Estimation des modèles à erreurs composées,* Annales de l'INSEE 7, pp. 43-72.

Mazodier, P., 1972. *l'Estimation des modèles à erreurs composées,* Annales de l'INSEE 7, pp. 43-72.
Mazodier, P. (red.), 1978. *The econometrics of panel data,* Speciaal nummer van Annales de l'INSEE 30-31.
Mundlak, Y., 1961. *Empirical production function free of management bias,* Journal of Farm Economics 43, pp. 44-56.
Mundlak, Y., 1978. *On the pooling of time series and cross section data,* Econometrica 46, pp. 69-86.
Nickell, S., 1981. *Biases in dynamic models with fixed effects,* Econometrica 49, pp. 1417-1426.
Scheffé, H., 1959. *The analysis of variance,* Wiley, New York.
Searle, S.R., 1971. *Linear models,* Wiley, New York.
Swamy, P.A.V.B., 1971. *Statistical inference in random coefficients regression models,* Springer, Berlijn.
Taub, A.J., 1979. *Prediction in the context of the variance-components model,* Journal of Econometrics 10, pp. 103-107.
Theil, H., 1971. *Principles of econometrics,* North-Holland, Amsterdam.
Van de Stadt, H. en T.J. Wansbeek, 1982. *Het effect van selectieve non-respons op de resultaten van een regressie-analyse,* Kwantitatieve Methoden 7, pp. 109-124.
Wansbeek, T.J., 1980. *Quantitative effects in panel data modelling,* Proefschrift, Rijksuniversiteit te Leiden.
Wansbeek, T.J. en A. Kapteyn, 1978. *The separation of individual variation and systematic change in the analysis of panel survey data,* Annales de l'INSEE 30-31, pp. 659-680.
Wansbeek, T.J. en A. Kapteyn, 1981. *Maximum likelihood estimation in incomplete panels,* intern CBS-rapport, Centraal Bureau voor de Statistiek, Voorburg.
Wansbeek, T.J. en A. Kapteyn, 1982. *A class of decompositions of the variance-covariance matrix of a generalized error components model,* Econometrica 50, pp. 713-724.

14. Steviger meten met een latent model

I.W. MOLENAAR

SAMENVATTING

In de gedrags- en maatschappijwetenschappen is het meten niet gemakkelijk: zodra menselijk gedrag in het geding is wordt het lastig om waarneming op bevredigende wijze te reduceren tot toekenning van getallen. Door deze toekenning te baseren op een meetmodel, dat geobserveerd gedrag relateert aan een onderliggende theorie, wordt duidelijker zichtbaar of de verkregen getallen inderdaad een zinvolle meting voorstellen.

Deze gedachte wordt uitgewerkt aan de hand van latente-trekmodellen voor ééndimensionale meting van attituden; de modellen van Guttman, Mokken en Rasch worden in dit licht besproken. Daarna wordt een algemene evaluatie geschetst van de genoemde modellen die met klassieke factoranalyse en testtheorie worden vergeleken. Tot slot volgt een toepassing op attituden van leerlingen ten opzichte van hun klasgenoten.

1. INLEIDING: METEN IS NIET GEMAKKELIJK

'Wat heet wetenschap' werd door de Nederlandse bewerkers van Chalmers (1976) gekozen als pakkende vertaling van 'What is this thing called science?' Uit de ondertitel blijkt dat het boekje gaat 'over aard en status van de wetenschap en haar methoden'; globaal weergegeven is het een inleiding over de denkbeelden die uit onvrede met het inductivisme werden ontwikkeld door Popper, Lakatos, Kuhn en enkele anderen. Ik vond na lezing mijn vooroordelen bevestigd dat wetenschapsfilosofen het voortdurend over Galileï en Einstein hebben, of in elk geval over de rol van empirie en theorie bij een paar grote ontdekkingen in de sterrenkunde en in de natuurkunde. Ook het grote publiek denkt bij wetenschap aan een laboratorium of een sterrenwacht. Maatschappijwetenschappen en gedragswetenschappen stemmen niet met dit bêta-beeld overeen, en worden in de publieke opinie met nogal wat wantrouwen bejegend: is dat wel 'echte' wetenschap?

Men kan zich afvragen of die 'zachtere' disciplines over minder groei van onomstreden kennis beschikken door hun jongere leeftijd, door hun maatschappelijk-economische rol of zelfs door een gemiddeld lagere intelligentie van hun beoefenaren. Hoewel ik persoonlijk een tot gamma afgedwaalde bêta ben, lijkt het mij volstrekt zinloos de bloei van de gamma-wetenschappen te willen bevorderen door er een natuurwetenschappelijk stempel op te drukken: het object van onderzoek is wezenlijk anders, en de methoden van onderzoek dienen daarmee rekening te houden. Deze lange inleiding heeft alleen ten doel een context te geven aan mijn centrale thema 'meten is niet gemakkelijk': zodra menselijk gedrag in het geding is wordt het erg lastig om waarneming te reduceren tot toekenning van getallen. Soms is het daarom verstandig van getaltoekenning af te zien en verdient verbale en impressionistische beschrijving de voorkeur. In andere gevallen zou dit evenwel een te groot verlies aan compac-

te, verifieerbare en overdraagbare informatie betekenen. Het gaat er dan om de problemen bij de meting zo zorgvuldig mogelijk te behandelen en daarmee de bij de reductie tot getallen opgelopen schade zo beperkt mogelijk te houden.

Dit artikel beoogt globaal na te gaan hoe de meting van mensen, en in het bijzonder van menselijke attituden, kan worden verbeterd door gebruik te maken van meettheorie en meetmodellen. Zekere technische aspecten van het begrip 'kwaliteit van de meting' zullen aan de orde komen. Daarbij mag niet worden vergeten dat die kwaliteit ook in sterke mate bepaald wordt door inhoudelijke vragen omtrent het begrip dat men bedoelt te meten en door de zorgvuldigheid waarmee de te meten persoon of groep van personen gestimuleerd wordt om welgeïnformeerde, eerlijke en betrouwbare antwoorden te geven. Een meting die daar geen aandacht aan schenkt is gedoemd te falen, ook als er een mooi wiskundig schaalmodel is gebruikt. Hoewel inhoudelijke tekortkomingen van dit type soms via de wiskundige analyse aan het licht komen, is dit zeker geen regel.

2. MEETMODEL ALS ZELFCONTROLEREND INSTRUMENT

'Measurement is the assigning of numbers to individuals in a systematic way as a means of representing properties of the individuals. Numbers are assigned to individuals according to a carefully prescribed repeatable procedure'. Bij dit citaat uit Allen & Yen (1979) valt allereerst op dat de getallen eigenschappen 'representeren'. Dit wordt in de meettheorie gepreciseerd tot de eis dat bepaalde relaties tussen de toegekende getallen corresponderen met empirische relaties tussen de gemeten objecten (hier personen, maar dat is niet essentieel). Bij nominaal meetniveau (rugnummers) geldt dat alleen voor de ongelijkheidsrelatie, bij ordinaal niveau ook de ordening, bij ratio-meting de verhouding, enz. Het is een probleem dat er bij veel eigenschappen in de gamma-wetenschappen maar weinig mogelijkheden zijn om empirische relaties van deze aard direct vast te stellen. Toch is het ook dan wenselijk de meetprocedure zo in te richten dat er mogelijkheden ontstaan om de kwaliteit van de getaltoekenning te onderzoeken en waar nodig te verbeteren. Daardoor ontsnapt de onderzoeker aan het gevaar van fiatmeting en extreem operationalisme ('deze test meet politieke progressiviteit omdat ik dat zeg' of 'intelligentie is wat deze test meet'), waarbij de meting bij voorbaat onaantastbaar wordt verklaard en haar tekortkomingen niet voor onderzoek vatbaar zijn.

Een meetmodel is kort gezegd een soort theorie van de meetprocedure, die vastlegt hoe het manifeste gedrag van de te meten persoon is gerelateerd aan een verondersteld onderliggend proces dat *latent* (niet direct observeerbaar) is. Daarbij moet het meetmodel zo specifiek zijn dat,

a. mogelijke statistische parameters van het onderliggende proces kunnen worden geschat;
b. er zekere restricties voor het manifest gedrag uit voortvloeien, die empirisch kunnen worden getoetst.

Wordt het meetmodel bij deze toetsing weersproken, dan was het een verkeerde theorie en beschikt de onderzoeker dus niet over metingen. Is het manifest gedrag conform aan de restricties afgeleid uit het meetmodel, dan be-

schikt de onderzoeker over zowel een theorie als metingen en als de toetsing voldoende streng was (waarover later meer) dan kan in beide vertrouwen worden gesteld. Merk op dat de *meting* in deze conceptie een *toegift van de theorie is:* pas als de theorie klopt zijn er metingen die de naam metingen waard zijn.

Aan de hand van een voorbeeld zullen wij de verschillende stappen bij de confrontatie van meetmodel en waarnemingen in het volgende toelichten. Voor een uitvoeriger uitleg raadplege men bijvoorbeeld Dawes (1972).

3. ATTITUDEMETING VOLGENS GUTTMAN

'Attitudes can be measured' was de triomfantelijke titel van een artikel van Thurstone (1928), waarin hij een meetmodel voorstelde dat via paarsgewijze vergelijking van stimuli tot schaling van stimuli op een continuüm voor een latente attitude leidde. De toekenning van schaalwaarden aan respondenten is evenwel niet in het meetmodel opgenomen. Toch wordt de techniek nog in talrijke inleidingen over schaalconstructie besproken, hoewel de dataverzameling vrij moeizaam is en de meetmethode zelden meer voor attitudemeting wordt toegepast. Hier kiezen wij daarentegen de Guttman-schaal als voorbeeld van een meetmodel, eveneens verouderd, maar ten eerste eenvoudiger en ten tweede nauwer verwant aan de hierna te bespreken andere latente-trekmodellen. In Molenaar (1982a) heb ik de Guttman-procedure vergeleken met het sorteren van aardappels door deze over een serie zeven met toenemende maaswijdte te leiden: de kleinste vallen al door de eerste zeef, de iets grotere door de tweede, enzovoorts totdat de grootste aardappels zelfs op de grofste zeef blijven liggen. Met een te meten persoon als aardappel en een zeef als een met eens/oneens te beantwoorden uitspraak is de analogie alleen verstoord doordat de items niet in opklimmende volgorde plegen te worden aangeboden en elke persoon elk item beantwoordt (een eenmaal gevallen aardappel bereikt de grovere zeef niet).

Het meetmodel van Guttman veronderstelt dat elk item i resp. elke persoon v correspondeert met één vaste positie ξ_i resp. ξ_v op een verondersteld latent continuüm voor de te meten eigenschap. Wanneer $\xi_v \geqslant \xi_i$ is persoon v het met item i eens (score 1), wanneer $\xi_v < \xi_i$ oneens (score 0). De schatting van personen én items bestaat in dit geval uit de gemeenschappelijke ordening van personen en items. De restricties van het meetmodel houden in dat iemand die het met totaal Y van de items eens is, volgens het model precies de Y gemakkelijkste items met eens en de overige met oneens beantwoordt: een respondent met een ander scorepatroon weerspreekt het model. De toetsingsprocedure is dus erg eenvoudig: orden alle k items naar dalend aantal personen met het eens-antwoord, dan moet elke persoon eerst een aantal enen en vanaf een zeker item uitsluitend nullen vertonen. Van de 2^k mogelijke antwoordpatronen zijn er dus slechts $k+1$ volgens het model toelaatbaar. Omdat het model deterministisch is, zijn uitzonderingen niet toegestaan.

4. LATENTE-TREKMODELLEN

In de algemene inleiding op het begrip *latente trek* bij Lord & Novick (1968, par. 16.1) staat o.m.: 'Any theory of latent traits supposes that an individual's

behavior can be accounted for, to a substantial degree, by defining certain human characteristics called *traits,* quantitatively estimating the individual's standing on each of these traits, and then using the numerical values obtained to predict or explain performance in relevant situations (...). Much of psychogological theory is based on a trait orientation, but nowhere is there any necessary implication that traits exists in any physical or physiological sense. It is sufficient that a person behaves as if he were in possession of an amount of each of a number of relevant traits and that he behaves as if these amounts substantially determined his behavior (...) The psychologist who wishes to use an examinee's responses to a mental test to make inferences about his psychological traits must have some knowledge of how psychological traits determine or are related to examinee's responses.

De trek wordt *latent* genoemd omdat zij niet direct observeerbaar (verborgen) is; het is een algemeen en platonisch begrip dat zich alleen indirect manifesteert in de concrete reactie op specifieke items. In de meeste modellen over de relatie tussen trekken en items postuleert men *lokale onafhankelijkheid:* in de deelverzameling van personen die op elke relevante trek dezelfde waarde hebben is de kansverdeling van de antwoorden tussen items stochastisch onafhankelijk; zij wordt alleen veroorzaakt door toevallige storingen. Die deelverzameling is, door het latente karakter van de trekken, uiteraard nooit in concreto te bepalen. In de verzameling van alle personen is er uiteraard juist een *positieve afhankelijkheid* tussen de items: weet ik dat iemand al enige items positief heeft beantwoord, dan zal hij/zij wel een gunstige positie op de relevante latente trek(ken) hebben en verwacht ik dus ook relatief veel positieve antwoorden op de overige items.

De assumptie dat een test door slechts één latente trek wordt bepaald (*unidimensionaliteit*) is nu nogal ingrijpend: geen enkele andere eigenschap van de te meten personen mag nog systematische invloed op de antwoorden uitoefenen. Toch streeft men naar een dergelijke eerlijke en ongecontamineerde meting, o.m. omdat het onethisch is als een test naar bv. geslacht of ras zou discrimineren, of omdat het resultaat op een test voor rekenvaardigheid niet door taalkennis of leessnelheid zou moeten worden beïnvloed. Daarmee is alleen bedoeld dat personen van *gelijke vaardigheid* (ongeacht hun ras, sexe, taalkennis of leessnelheid) ook dezelfde score behoren te krijgen; het is dus best mogelijk dat die vaardigheid ongelijk verdeeld is naar ras, sexe, enz., maar de itemkeuze mag niet ten nadele van zo'n groep werken door een specifieke interactie (zie ook Van der Flier, 1980). McNemar schreef al in 1946: 'measurement implies that one characteristic at a time is being quantified'. Hierbij moet overigens wel worden opgemerkt dat weglating van items die niet strikt aan deze eis voldoen, kan leiden tot een ongewenste versmalling van het te meten begrip. Deze bandbreedte-vernauwing is trouwens een gevaar bij elke ééndimensionale schaalconstructie die te veel op de wiskundige betrouwbaarheid en te weinig op de inhoud let: een verzameling items waarin telkens dezelfde vraag in iets andere bewoordingen wordt herhaald is doorgaans erg unidimensioneel, en betrouwbaar, maar weinig valide. Wanneer we in de woorden van Lord (1980) een test opvatten als 'a device of obtaining a sample of behavior', wensen we dat deze gedragssteekproef representatief is voor het bedoelde gebruik van de test.

In de volgende paragrafen worden een aantal modellen besproken voor tests bestaande uit *dichotome* items (twee antwoordcategorieën, bv. goed/fout of eens/oneens) die één latente trek meten en daarin *cumulatief* zijn verondersteld (de kans op een positief antwoord is voor elk item een niet-dalende functie van de waarde op de latente trek).

Uit onze assumptie volgt, dat de kans op het ja-antwoord voor elk item i een niet-dalende functie $f_i(\xi)$ van de latente eigenschap is. Men noemt $f_i(\xi)$ de *item-karakteristieke curve* (Engels *traceline*). De diverse in de volgende paragrafen te bespreken latente-trekmodellen onderscheiden zich onderling door de assumpties die over $f_i(\xi)$ worden gemaakt. Deze zijn niet direct empirisch verifieerbaar omdat ξ latent is, maar zij hebben wel observeerbare consequenties, die toelaten iets over hun aannemelijkheid te zeggen.

Binnen een specifiek latente-trekmodel probeert men nu vragen te beantwoorden zoals:

1. Wat zijn de beste schattingen voor de parameters van de items en van de personen?
2. Vormt een gegeven item verzameling een (goede) schaal in de zin van het model?
3. Kan ik uit een gegeven itemverzameling door stapsgewijs opbouwen of stapsgewijs weglaten een goede schaal verkrijgen?

5. SCHAALANALYSE VOLGENS MOKKEN

Mokken (1971) introduceerde een probabilistische variant van de Guttman-schaal: een negatief antwoord op een item met een positief antwoord op een moeilijker item is nu niet meer strikt verboden, maar behoort zich slechts zelden voor te doen. Mokken drukt de mate van modelconformiteit uit via de H-coëfficiënt van Loevinger; voor elk paar items geldt $H=1$ als de items geen uitzonderingen op de Guttman-regel vertonen en $H=0$ als ze statistisch volstrekt onafhankelijk van elkaar worden beantwoord.

Voor een uitvoerige behandeling van de Mokken-schaal raadplege men naast Mokken (1971) bijvoorbeeld STAP User's Manual (1980), Stokman & Van Schuur (1980) en Molenaar (1982b,c), waarin de kritiek van Jansen (1982) wordt beantwoord. Illustraties van de itemcurven en toelichting op de H-coëfficiënt vindt men behalve in de genoemde bronnen ook in Molenaar (1982a). Hier zal alleen kort worden aangegeven hoe ook in dit geval het meetmodel dient om de kwaliteit van de meting te beschermen.

De assumpties van het model zijn dat een aselecte steekproef van n personen antwoord geeft op k dichotome items met monotone itemkarakteristieke curven, die elkaar niet snijden en zodanig verlopen dat de genoemde H-coëfficiënten hoog genoeg zijn. Uit deze eisen volgen restricties voor de geobserveerde antwoordpatronen, minder streng dan die van Guttman maar voldoende om tot een ordening van personen en items te geraken op grond van het totaal aantal positieve antwoorden.

Omdat het model probabilistisch is, zijn die restricties niet absoluut maar relatief: bepaalde patronen mogen niet te vaak voorkomen, terwijl zij in het deterministische Guttman-model volstrekt niet mochten voorkomen. De monotonie-assumpties worden in dit onderzoek via de H-coëfficiënt hoogstens

indirect op de proef gesteld; voor voorstellen voor aanvullend onderzoek raadplege men Molenaar (1982c).

Naast een toets voor modelconformiteit ontwikkelde Mokken onder meer procedures voor de stapsgewijze selectie van modelconforme items uit een grotere itemverzameling (de punten 2 en 3 uit het slot van paragraaf 4). Merk op dat de in punt 1 aldaar genoemde schatting van persoons- en itemparameters zich hier tot een ordening beperkt, omdat Mokken slechts verdelingsvrije eisen formuleert voor de vorm van de itemcurven. Zoals nader toegelicht in Molenaar (1982c), zijn de modeleisen deels afhankelijk van de frequentieverdeling van de te meten groep personen op de latente trek: het onderzoek naar de schaalbaarheid dient dan ook plaats te vinden aan een groep personen die representatief is voor de populatie waarop men de schaal wil gebruiken.

6. SCHAALANALYSE VOLGENS RASCH

Rasch (1960) voegt aan de assumpties van dichotome, unidimensionele, cumulatieve en locaal onafhankelijke items nog toe dat de totaalscore afdoende (sufficiënt) is voor de latente trek ξ. Dan kan worden bewezen dat de itemcurven alle evenwijdig en logistisch zijn:

$$P\text{ (persoon } v \text{ antwoordt positief op item } i) = e^{\xi_v - \sigma_i}/(1 + e^{\xi_v - \sigma_i}),$$

waarbij de itemparameter σ_i de ξ-waarde voorstelt waarbij precies de helft van de personen het item positief beantwoordt.

Voor twee items i en j geldt nu, in de verkorte notatie $\theta_v = e^{\xi_v}$ en $\varepsilon_i = e^{-\sigma_i}$,

$$\frac{P\text{ (persoon } v\ i \text{ goed en } j \text{ fout)}}{P\text{ (persoon } v\ i \text{ fout en } j \text{ goed)}} = \frac{\dfrac{\theta_v \varepsilon_i}{1+\theta_v \varepsilon_i} \cdot \dfrac{1}{1+\theta_v \varepsilon_j}}{\dfrac{1}{1+\theta_v \varepsilon_i} \cdot \dfrac{\theta_v \varepsilon_j}{1+\theta_v \varepsilon_j}} =$$

$$= \frac{\theta_v \varepsilon_i}{\theta_v \varepsilon_j} = \frac{\varepsilon_i}{\varepsilon_j} = e^{-\sigma_i + \sigma_j}$$

Het quotiënt hangt niet van θ_v af, waaruit men kan afleiden dat de bepaling van het verschil $\sigma_i - \sigma_j$ tussen twee itemparameters voor elke deelgroep van personen tot hetzelfde resultaat leidt, afgezien van steekproef-onnauwkeurigheid. Dit geldt ook voor verschillen $\xi_v - \xi_w$ tussen personen bij deelgroepen van items. Deze zg. specifieke objectiviteit, het ontbreken van interactie tussen items en personen, is vanuit meettheoretisch oogpunt uiterst wenselijk. Men kan hierdoor de scores van personen die verschillende itemdeelgroepen hebben beantwoord, achteraf vergelijkbaar maken. Zoals kort beschreven door Molenaar (1982a) en uitvoerig door Wright & Stone (1979) en Fischer (1974), opent dit de weg naar allerlei interessante toepassingen. De prijs daarvoor is dat het meetmodel van Rasch strenge eisen stelt, zodat het construeren van Rasch-schalen niet eenvoudig is. De toetsing van de Rasch-assumpties wordt o.m. uitvoerig besproken door Van den Wollenberg (1979, 1982).

Nog sterker dan bij Mokken geldt dat een directe vertaling van de assump-

ties in observeerbare consequenties met problemen gepaard gaat, zodat het lastig is om ondubbelzinnig vast te stellen of geobserveerde data aan het Rasch-model voldoen. Er zijn mede daarom geen pasklare algorithmen om uit een gegeven itemverzameling door stapsgewijze opbouw of stapsgewijze eliminatie een Rasch-conforme deelschaal te selecteren.

Het is een groot voordeel van het Rasch-model dat de schatting van de itemparameters σ_i en de persoonsparameters ξ_v een statistisch welomschreven probleem is, zodat de gebruiker de betrouwbaarheid van deze schattingen kan berekenen. Het algorithme voor de schatting is gecompliceerd, maar er zijn diverse computerprogramma's beschikbaar, waarvan bijvoorbeeld PML van Gustafsson (1979) zowel flexibel als nauwkeurig en snel mag worden genoemd.

Een variant op het Rasch-model is dat van Birnbaum, waarin de itemcurven verschillende hellingen mogen hebben. Bij toepassing op meerkeuze-items in vorderingen toetsen kan bovendien nog een gisparameter per item worden ingebouwd. In deze varianten is het totaal aantal goede antwoorden niet langer afdoende, en is de parameterschatting gecompliceerder. Dit model wordt uitvoerig besproken door Lord (1980); de parameterschatting met zijn programma LOGIST vereist overigens zeer grote steekproeven. Andersen (1982) gaat in op varianten voor items met ongelijke helling en items met meer dan twee antwoordcategorieën, waarbij de hellingen resp. categorie-afstanden tevoren bekend moeten zijn. Gelukkig correleert de via deze varianten aan een persoon toegekende score in het algemeen vrij hoog met de simpele dichotome somscore, zodat het Rasch-model bij afwezigheid van gissen tamelijk robuust lijkt te zijn (zie bv. Forsyth et al., 1981).

7. EVALUATIE EN DISCUSSIE

In Molenaar (1982a) is een poging ondernomen om de hier in vogelvlucht besproken latente-trekmodellen vanuit consumentenstandpunt te vergelijken, zowel onderling als met de klassieke methoden van Thurstone-Chaves, Likert, factoranalyse en testtheorie (Cronbach's alfa). Er blijken dan wezenlijke verschillen te bestaan naar meetmodel c.q. naar kwaliteitscriterium voor schaalbaarheid. Toch rijst de vraag of de uiteenlopende schaalmethoden niet grotendeels dezelfde resultaten opleveren. In een specifieke toepassing op items over werkoriëntaties, met aandacht voor allerlei beslispunten voor de gebruiker, bleken diverse schaaltechnieken veelal tot ongeveer gelijkluidende eindresultaten te leiden (De Vos, 1980), al waren er uitzonderingen en al was het datamateriaal op een aantal punten niet ideaal.

Een algemeen antwoord op de vraag naar de beste schaaltechniek lijkt mij niet mogelijk, omdat het doel van de meting, de aard en omvang van het databestand, de beschikbaarheid van tijd en van programmatuur en enkele soortgelijke onderzoek-specifieke factoren bij de keuze van techniek en model van doorslaggevende invloed zullen zijn. Niettemin zal hieronder worden getracht een paar algemene opmerkingen te maken over het keuzeproces. De ervaringen bij de consultatie door onderzoekers hebben mij er toe verleid daarbij ook enige zaken op te nemen die, eenmaal gezegd, nagenoeg vanzelfsprekend zijn, maar waaraan toch te vaak wordt voorbijgegaan in de onderzoekpraktijk.

Bestaande meetinstrumenten

Vanuit het principe 'eerst kennis nemen van de bestaande wetenschappelijke literatuur' zou ieder die een eigen schaal gaat construeren, eerst moeten nagaan welke schalen voor het beoogde begrip er al bestaan. Binnen de testpsychologie is men ver gevorderd met de screening en codificatie van tests (Algera et al., 1978, resp. Informatie- en Documentatiecentrum van het Ned. Instituut voor Psychologie, p/a mw. J.C. van Vliet-Mulder, Batavierenweg 15, 6616 AR Hernen); veel data-materiaal berust bij het Steinmetz-Archief, Herengracht 410, 1017 BV Amsterdam. Ook wanneer een bestaande schaal niet zonder verandering bruikbaar zou zijn, kan de ervaring die anderen met bestaande items hebben opgedaan van veel nut zijn.

Aantal antwoordcategorieën

Het aanbieden van te veel of te weinig categorieën leidt tot frustratie bij de respondenten. Bij correlationeel onderzoek (item- en factoranalyse) wordt de equidistantie van de categorieën zonder nader onderzoek aangenomen, en als dat klopt geldt voor de verwerking 'hoe meer categorieën hoe beter'. Exploratie van optimale schaalwaarden voor de categorieën is mogelijk via de door Gifi (1981) beschreven algorithmen, die evenwel voor item-categorieën met een lage frequentie nogal instabiele waarden opleveren (kanskapitalisatie). De in de voorafgaande paragrafen beschreven latente-trekmodellen zijn grotendeels alleen voor dichotome items bruikbaar. De onderzoeker moet dan kiezen tussen dichotome aanbieding aan de respondent, met de risico's van frustratie en van te scheve verdelingen, of dichotomisering achteraf, met het risico dat die beslissing te kort schiet door kanskapitalisering of door extremiteitstendenties bij sommige respondenten.

Verdeling van itempopulariteiten en van scores

Zowel bij de keuze van items als bij de keuze van dichotomisering binnen items verdient de frequentieverdeling van scores in de totale respondentengroep meer aandacht dan daaraan doorgaans wordt besteed. Dit geldt eveneens voor de verdeling van de itempopulariteiten (d.w.z. item p-waarden, de fractie personen met een positief antwoord). Lord (1980) signaleert de voordelen van een itemkeuze die de schaal onderscheidend maakt in het gebied waar men de respondenten ook echt wenst te onderscheiden. Items met een populariteit dicht bij ½ zijn optimaal, in de zin van informatiegehalte, voor de doorsnee-respondent, maar niet voor de selectie van bijvoorbeeld de top tien procent of de laagste vijf procent. Een kosten/baten-analyse zal er dikwijls toe leiden om na een vooronderzoek items te verwijderen met extreem hoge of extreem lage populariteit: ook als zij sterk met de latente trek zouden samenhangen vormen zij voor de meeste respondenten een bijna overbodig onderdeel van de schaal. Toch wordt daarmee ook een controlemogelijkheid verwijderd. De ervaring leert dat klassieke itemanalyse doorgaans uitloopt op veel items van middelmatige populariteit (in de hoop dat de invloed van meetfouten door uitmiddelen wordt verminderd), terwijl Guttman- en Mokken-procedures van-

uit de assumptie dat elk item op één plaats goed discrimineert, doorgaans meer gespreide populariteiten opleveren, waarbij de respondenten door de items in equivalentieklassen worden gesorteerd. Binnen het Rasch-model geldt dat veel minder, omdat ook zeer matig discriminerende items aan dat model kunnen voldoen (Molenaar 1982b, c).

Assumpties van het schaalmodel

Elk schaalmodel kent zijn eigen assumpties die sterk of zwak, talrijk of schaars kunnen zijn. Het lijkt mij verstandig in het licht van de data en van de bedoeling van het onderzoek zorgvuldig te evalueren welke assumpties het betreft, of deze in het concrete geval voor wenselijk worden gehouden en of het plausibel is dat de data eraan voldoen. Van eventuele schendingen kan men vragen of (en hoe) deze kunnen worden aangetoond, en of hun gevolgen voor de uiteindelijke meting op hun ernst kunnen worden onderzocht. Te vaak vaart men blind op één index of één significantietoets voor de passing van het model. De robuustheid en nauwkeurigheid van die index, het onderscheidingsvermogen van de toets en de robuustheid van de meting tegen assumptieschending zijn in veel gevallen onvoldoende bekend. Wel groeit het methodologisch onderzoek naar deze vragen vrij snel, in harmonie met een algemene tendentie in de statistiek om over de grenzen van één wiskundig model heen te kijken (Molenaar, 1979).

Undimensionaliteit

Al in paragraaf 4 is gesignaleerd dat overdrijving van deze eis tot absurde vernauwing van het te meten begrip kan leiden. Het gebruik van veel items of veel indicatoren is in de sociale wetenschappen juist populair geworden, omdat de bedoelde begrippen betrekkelijk breed zijn en zich niet door één vraag laten meten. Nu moet door zorgvuldige overweging van de inhoud van de items worden nagegaan of verwijdering van de afwijkende items wel met de validiteitseis verenigbaar is. Waar er duidelijk enkele deelschalen in de data kunnen worden aangetoond, kan men de subjecten misschien beter per deelbegrip een score toekennen, maar de eis van spaarzame beschrijving moet niet onder de tafel geraken en de scores per deelschaal zullen doorgaans minder betrouwbaar zijn door het korter zijn van die schaal. Het aardige van Cronbach's alfa als kwaliteitsindex is dat zowel de testlengte als de heterogeniteit erin wordt verdisconteerd; wie begint met een itemverzameling te beperken ziet alfa doorgaans eerst stijgen en dan weer dalen. Bij de modellen van Mokken en Rasch moet de testlengte, en daarmee de nauwkeurigheid van de meting, afzonderlijk in de overwegingen worden betrokken, hetgeen bij Mokken eenvoudiger is dan bij Rasch.

Voordeel latente-trekmodellen

De latente trek-modellen gaan ervan uit dat de te meten eigenschap een continue kansverdeling heeft en los van de gebruikte items bestaat; met het discrete karakter van die items wordt expliciet rekening gehouden. Dat ondervangt

minstens een deel van de bezwaren tegen de klassieke testtheorie en de factoranalyse, waarvan de assumpties voor een beperkt aantal discrete items al bij voorbaat onjuist moeten zijn. Bovendien zijn de assumpties in de nieuwe modellen beter zichtbaar en stringenter: de onderzoeker stelt zich kwetsbaarder op, maar kan dan ook de meting met meer vertrouwen gebruiken als de schaal de beproeving doorstaat. Een schaalprocedure 'waar altijd wel wat uit komt' is natuurlijk snel en gemakkelijk in het gebruik, maar slaat bij misbruik geen alarm, zodat misschien met ontoereikende metingen verder wordt gewerkt.

8. EEN VOORBEELD: LEERLINGATTITUDEN T.A.V. KLASGENOTEN

In het kader van een onderzoek naar optimale schoolgrootte heeft Stoel (1980) een schoolbelevingsschaal ontwikkeld: een proefversie van 78 items werd voorgelegd aan 685 leerlingen van vijf scholen, waarna de definitieve versie werd gebruikt bij de 7.016 leerlingen in het hoofdonderzoek. Onder verwijzing naar Stoel (1980, deel II) voor de inhoud en de constructieprocedure, gebruiken wij de scores van 679 leerlingen uit het vooronderzoek op vijf items als illustratie van het betoog over meetmodellen. Deze items betreffen de relatie van de leerling met zijn/haar klasgenoten en vertonen maar weinig samenhang met de overige aspecten van schoolbeleving (met plezier naar school gaan, persoonlijke en professionele waardering voor de docenten, oordeel over eigen prestaties en schoolorganisatie).

Bij de vijf onderstaande uitspraakteksten (in het volgende door een steekwoord aangeduid) moest elke leerling één van de vier antwoorden (bijna altijd; vaak; soms; bijna nooit) omcirkelen. Bij de met R gemerkte items hebben wij de scoring 1 t/m 4 omgekeerd, zodat een hoge score steeds een positieve attitude aangeeft.

AARDIG – R –	ik heb het gevoel dat mijn klasgenoten mij wel aardig vinden (4= bijna altijd).
STOM –	ik denk dat veel van mijn klasgenoten mij nogal stom vinden (4= bijna nooit).
HEKEL –	ik denk dat de leerlingen uit de klas een hekel aan mij hebben (4= bijna nooit).
EENZAAM –	ik voel me eenzaam op school (4= bijna nooit).
OPSCHIETEN – R	ik vind dat ik met mijn klasgenoten goed kan opschieten (4= bijna altijd).

Merk op dat 'stom' in modern scholierenjargon eerder op sociale acceptatie dan op intellectuele prestaties slaat; het item vertoont geen verband met de hier niet besproken subschaal die het zelfoordeel over deze prestaties meet.

Tabel 1 rapporteert een klassieke itemanalyse: Cronbach's alfa (een ondergrens voor de testbetrouwbaarheid) is 0,68 en zou dalen als een item werd verwijderd. In het hoofdonderzoek heeft Stoel de schaal tot elf items uitgebreid, waarbij alfa tot 0,79 stijgt.

De items correleren onderling matig; men kan berekenen dat de laagste correlatie (0,21 tussen AARDIG en EENZAAM) door de sterk verschillende mar-

ginale frequenties van de items als hoogst bereikbaar maximum niet 1,00 heeft maar 0,65. Wel heeft het EENZAAM-item over de hele linie de zwakste samenhang met de rest van de schaal.

Tabel 1. Itemanalyse bij scoring 1, 2, 3, 4 met $\alpha_{(i)}$ bij weglating van i.

item *i*	$r_{i,\ tot}$	$r_{i,\ rest}$	$\alpha_{(i)}$	STOM	HEKEL	EENZ.	OPSCH.
AARDIG (R)	0,72	0,44	0,64	0,31	0,34	0,21	0,37
STOM	0,69	0,50	0,60	-	0,55	0,32	0,27
HEKEL	0,69	0,53	0,60		-	0,24	0,36
EENZAAM	0,56	0,33	0,67			-	0,21
OPSCH. (R)	0,69	0,44	0,63				-
			α = 0,68 voor de schaal				

Tabel 2. Frequentieverdeling somscore (n = 679)

SOM	≤7	8	9	10	11	12	13	14	15	16	17	18	19	20
frekw.	0	4	3	3	7	6	28	55	73	65	113	116	131	75
cum.%	0	1	1	2	3	3	8	16	26	36	53	70	89	100

De somscore over de vijf items blijkt wegens de in tabel 3 weergegeven scheve verdeling per item nooit de theoretische minimumwaarde van 5 te bereiken. Hoewel de verdeling scheef is, neemt de som voldoende verschillende waarden aan (zie tabel 2) om de leerlingen naar hun algemene relatie tot de klasgenoten te onderscheiden.

Tot nu toe zijn de categorieën als 1, 2, 3, 4 gecodeerd. Met het programma HOMALS (Gifi, 1981) kan men een optimale codering van zowel itemcategorieën als persoonsscores bereiken, in de zin dat elke persoon een score krijgt die zo dicht mogelijk ligt bij de door hem/haar aangekruiste categorieën. Tabel 3 geeft per item per categorie zowel de HOMALS-score als de waargenomen frequentie.

Tabel 3. Frequentieverdeling per item en HOMALS categoriescores

	1	2	3	4	1	2	3	4
AARDIG	53	260	220	146	−1,38	−0,53	0,55	0,61
STOM	10	15	223	431	−2,49	−2,91	−0,66	0,50
HEKEL	2	13	221	443	−4,49	−3,00	−0,77	0,49
EENZAAM	10	29	129	511	−0,90	−2,03	−0,68	0,31
OPSCH.	14	86	209	370	−2,52	−1,32	−0,03	0,42

De categorieën 3 en 4 bij item AARDIG liggen dan dicht bij elkaar: of iemand daar 'bijna altijd' dan wel 'vaak' omcirkelt heeft dus weinig verband met de score op de andere items. Bij STOM en EENZAAM is de ordening verstoord voor categorieën 1 en 2. Door de berekening te herhalen voor gedeelten van het leerlingenbestand hebben wij gevonden dat alle scores die met frequenties lager dan 20 in tabel 3 optreden tamelijk instabiel zijn (de 2,49 voor STOM kan bijvoorbeeld tussen 1,9 en 2,6 schommelen), terwijl de frequent waarge-

Tabel 4. Mokken-schaal bij twee dichotomiseringskeuzen

positief	itempopulariteiten					H	alfa
3, 4	0,54	0,85	0,94	0,96	0,98	0,52	0,51
4	0,21	0,54	0,63	0,65	0,75	0,37	0,65

nomen scores hoogstens in de tweede decimaal veranderen. De instabiliteit is overigens geen afdoende verklaring voor de verstoorde ordening bij STOM en EENZAAM.

De in tabel 2 gebruikte somscore over de vijf met 1, 2, 3, 4 gecodeerde items blijkt 0,97 te correleren met de HOMALS-score. Toch kan iemand bij vaste HOMALS-score −0,2 als somscore 14, 15, 16 of 17 hebben en bij vaste somscore 14 loopt de HOMALS-score van −1,9 tot −0,2 uiteen. Voor de groep als geheel leidt de HOMALS-meting ondanks deze uitzonderingen tot ongeveer gelijkluidende conclusies.

Voor de bestaande Mokken- en Rasch-procedures moeten de antwoordcategorieën per item tot twee worden samengevoegd (zg. dichotomiseren), al is de schrijver van dit artikel bezig met een Mokken-model voor meer dan twee categorieën en bespreken Fischer (1974) en Andersen (1982) een soortgelijke uitbreiding voor het Rasch-model. Bij de vraag naar de verstandigste dichotomisering kan al direct uit tabel 3 worden opgemaakt dat samenvoegen van de antwoorden 2, 3 en 4 tot 'positief' tot extreem weinig negatieve antwoorden leidt. Dit geldt, hoewel minder, ook als 3, 4 = positief en 1, 2 = negatief wordt gekozen (tabel 3 en 5); de schaalcoëfficiënt *H* blijkt dan in tabel 4 wel vrij hoog. Kiest men alleen 4 als positief dan is de scoreverdeling gelijkmatiger en alfa is hoger, maar *H* is lager.

Er zijn dan meer respondenten die fouten maken tegen het perfecte schaalpatroon, d.w.z. gevallen waarin een gemakkelijk item negatief en een moeilijker item positief wordt beantwoord. Een telling van dit soort fouten per respondent is in tabel 5 weergegeven. Zijn de items naar opklimmende populariteit gerangschikt, dan is bijvoorbeeld bij score 3 het perfecte patroon − − + + +; iemand die daarentegen + − − + − antwoordt maakt vier fouten, want er zijn vier itemparen met + −.

Nader onderzoek van de dubbele monotonie, hier niet weergegeven, toont dat er bij de dichotomisering 4 = positief nogal veel verstoringen zijn. Bij 3, 4 = positief daarentegen is er niet alleen dubbele monotonie, maar ook vertonen 620 van 679 leerlingen een perfect scorepatroon, waarbij een totaalscore van *Y* inhoudt dat precies de makkelijkste *Y* items positief zijn beantwoord. Wel was dit voor de personen met score 5 vanzelfsprekend, maar tabel 5 toont ook voor de anderen weinig fouten. Deze inhoudelijke interpretatiemogelijkheid van een totale score achten wij een belangrijk onderdeel van een Mokkenschaal: men weet beter wat men meet als de getaltoekenning ook impliceert welke items positief zijn beantwoord.

Wanneer voor sommige items alleen 4 als positief wordt gescoord en voor andere 3 en 4, blijkt uit hier niet weergegeven berekeningen dat de eisen van een gelijkmatige scoreverdeling en van een hoge *H* elkaar enigszins tegenwerken. Het is bovendien gevaarlijk alle denkbare dichtomiseringen 'uit te probe-

Tabel 5. Frequentieverdeling van totale score (verticaal) en aantal fouten in het scorepatroon (horizontaal). Een mogelijk maar niet geobserveerd resultaat is met 0, een logisch onmogelijk resultaat met – aangeduid.

	3, 4 = pos. aantal fouten									4 = pos. aantal fouten							
score	0	1	2	3	4	5	6	totaal	score	0	1	2	3	4	5	6	totaal
0	4	–	–	–	–	–	–	4	0	49	–	–	–	–	–	–	49
1	3	0	2	1	1	–	–	7	1	49	13	11	23	2	–	–	98
2	4	3	2	0	1	1	0	11	2	18	27	54	13	6	1	1	120
3	68	18	3	2	2	0	0	93	3	81	26	21	18	11	3	0	160
4	202	12	6	4	1	–	–	225	4	135	23	5	4	10	–	–	177
5	339	–	–	–	–	–	–	339	5	75	–	–	–	–	–	–	75
tot.	620	33	13	7	5	1	0	679	tot.	407	89	91	58	29	4	1	679

ren', omdat in het hoofdonderzoek bij andere leerlingen gemakkelijk zou kunnen blijken dat de schijnbaar beste keuze dan toch weer tegenvalt, omdat zij te veel op een toevalligheid in de data van het vooronderzoek heeft voortgebouwd.

Het model van Rasch blijkt voor onze vijf items bij dichotomisering 4 = positief slecht te passen (item OPSCH. heeft een minder steile curve dan de andere items). Bij dichotomisering 3, 4 = positief geldt dit verschijnsel enigszins voor AARDIG, maar het is slechts op een scoregroep van 7 individuen gebaseerd en de algemene toets van Andersen wijst geen afwijkingen aan (P = 0,39; slechts twee groepen omdat de scores 1 en 2 maar 7 resp. 11 keer optreden). Een voordeel van het Rasch-model is dat de Rasch-scores per item en per persoon van een betrouwbaarheidsinterval zijn voorzien. Zoals bij vijf items en 679 personen te verwachten viel, kan de parameter voor itemmoeilijkheid nauwkeurig worden geschat, maar de score per persoon is erg onnauwkeurig. Tegenover de voordelen van specifieke objectieve meting staat in dit geval dus het nadeel dat de meting met slechts vijf gedichotomiseerde items grof is en tot een erg scheve scoreverdeling leidt.

De Rasch- en Mokken-scores bij de dichotomisering 3, 4 = positief correleren beide 0,88 met de oorspronkelijke somscore uit tabel 2. Wat dat betreft gaat er door de dichotomisering dus niet veel informatie verloren, al kan bijvoorbeeld een score van 3 op de gedichotomiseerde schaal een somscore 14, 15, 16, 17 of 18 betekenen. Voor Stoel's globale onderzoek naar de invloed van schoolgrootte op schoolwelbevinden zal de conclusie niet wezenlijk door de keuze van het schaalmodel worden beïnvloed. Bij een voor individuele advisering bestemde schaal zou voorzichtiger meting gewenst zijn. De uitbreiding tot elf items is daarom nuttig, maar in dit artikel hebben wij de versie van vijf items als illustratie gebruikt omdat die zich compacter in tabellen laat weergeven.

BIBLIOGRAFIE

Algera, J.A. et al., 1978. *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen.* Nederlands Instituut voor Psychologen.

Allen, M.J. & Yen, W.M. *Introduction to measurement theory,* Monterey: Brooks/Cole.

Andersen, E.B., May 1982, to be published. *A general latent structure model for contingency table data,* paper presented in Princeton.

Chalmers, A., 1981. *Wat heet wetenschap,* Boom: Meppel.

Dawes, R.M., 1972. *Fundamentals of attitude measurement,* Wiley: New York.

Fischer, G. H., 1974. *Einführung in die Theorie Psychologischer Tests,* Bern: Huber.

Flier, H. van der, 1980. *Vergelijkbaarheid van individuele testpresentaties,* Dissertatie V.U. Amsterdam, Lisse: Swets & Zeitlinger.

Forsyth, R., Saisangjan, U. & Gilmer, R., 1981. *Some empirical results related to the robustness of the Rasch-model,* Applied Psychological Measurement, 5, pp. 175-186.

Gifi, A., 1981. *Nonlinear multivariate analysis,* Afd. Datatheorie FSW, R.U. Leiden.

Gustafsson, J.E., 1979. *PML: A computer program for conditional estimation and testing in the Rasch-model for dichotomous items,* Reports from the Institute of Education, University of Göteborg, no. 85.

Jansen, P.G.W., Roskam, E.E.Ch.I., & Van den Wollenberg, A.L., 1982. *De Mokkenschaal gewogen,* Tijdschrift voor Onderwijsresearch, 7, pp. 31-42.

Lord, F.M., 1980. *Applications of item response theory to practical testing problems,* Erlbaum: Hillsdale N.J.